Computational studies of III-V nitride semiconductors

Martina Miskufova

December 19, 2011

A thesis submitted in fulfillment of the requirements of an EngD degree in the Chemistry department of University College London (UCL)

I, Martina Miskufova confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

GaN and related nitride compounds have found many applications in optoelectronic devices. Point defects introduce energy levels into the band gap and alter the electrical and optical properties of GaN. Previous studies focussed on studying point defects with density functional theory (DFT), periodic boundary conditions (PBC) and relatively inaccurate LDA energy functionals. We aim to improve on the deficiencies of this method by implementing a quantum mechanical/molecular mechanical (QM/MM) scheme, which has been specifically designed for the study of point defects; we use a hybrid functional and a formal charge scheme for the MM model. We offer an explanation for why p-type doping is difficult to achieve in GaN; the exothermic formation energies of the Ga interstitial and N vacancy at the VBM are thought to be the main cause. We suggest that the processes responsible for a variety of DLTS signals between 0.18-0.67eV below the CBM may be due to Ga interstitial 3+/2+, 2+/1+ transitions, N interstitials (1+/0) and Ga vacancies (2-/1-, 3-/2-). We attribute the ODMR signal indicating a deep donor state 0.7eV below the conduction band to the N interstitial 0/1- transition. Finally, our results support previous suggestions that Ga vacancies may be the cause of yellow luminescence in GaN.

Further refinements of the model, especially improving the basis set, are recommended in the future, as well as a more detailed investigation into the causes of discrepancy between our model and PBC calculations.

We use the MM model to study the properties of ternary alloys of AlN, GaN and InN, and to find their thermodynamically stable configurations. Our results are in good agreement with PBC DFT calculations. These structures are not observed experimentally; we suggest that this is a growth phenomenon. Our results also support previous findings that epitaxial strain stabilises highly internally strained alloys.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor, Prof. Richard Catlow for many useful scientific insights over the years. Dr. Alexei Sokol has contributed greatly to my understanding of QM/MM methods and defect systems in particular. Dr. Aron Walsh has been a valuable source of information about periodic boundary condition methods and defect energetics as well as giving expert advice on alloy modelling. Dr. Scott Woodley has been invaluable in helping with questions regarding GULP methodology, potential fitting and solid solutions. Dr.Ricardo Grau-Crespo has been very kind to discuss alloy modelling with me. Prof. Paul Sherwood and Dr. Thomas Keal have been very responsive to my pleas for technical help with ChemShell. Joanna Walsh and Dr. Tom Sadler have helped me understand the experimental side of nitride research. Without the patient help of all these people, I would never had got this far. On the financial side, I would like to acknowledge the EPSRC and my sponsor company Accelrys for their generosity and the opportunities they have given me to see how research finds its applications in industry. At Accelrys, I would particularly like to mention Dr. Paul Strodel, who with the patience of a saint answered all of my technical questions. The calculations in the QM/MM section would not have been possible without the national supercomputing facility, HeCToR, and I have been very lucky to be allocated time on this expensive machine to perform my defect calculations.
Finally, I would like to thank Goodenough College for providing accommodation and a wonderful community for four years, my friends, especially Dr. Amina Aitsi-Selmi, Katherine Bash and Craig MacKenzie, for all their support and encouragement during the difficult times, and, last but not least, my wonderful family, particularly my parents for always being there for me, and my 8-year-old sister, Simonka, for asking important questions, such as what is a doctorate.

Abbreviations

CBM Conduction band minimum
DLTS Deep level transient spectroscopy
ECP Effective core potential
GGA Generalised gradient approximation
HOMO Highest occupied molecular orbital
IP Ionisation potential
LDA Local density approximation
LSDA Local spin density approximation
LUMO Lowest unoccupied molecular orbital
MM Molecular mechanics
PAS Positron annihilation spectroscopy
QM Quantum mechanics
SCF Self-consistent field
VBM Valence band maximum
YL Yellow luminescence

Publications

1. On the problem of cluster structure diversity and the value of data mining, A. A. Sokol, C. R.A. Catlow, M.Miskufova, S.A. Shevlin, A.A.Al-Sunaidi, A.Walsh, S. M.Woodley, Phys Chem Chem Phys 30 **8438 - 8445** (2010)

2. Advances in computational studies of energy materials, C. R. A. Catlow, Z. X. Guo, M. Miskufova, S. A. Shevlin, A. G. H. Smith, A. A. Sokol, A. Walsh, D. J. Wilson, S. M. Woodley, Phil.Trans.R. Soc. A 28 **3379 - 3456** (2010)

3. Electron and hole stability in GaN and ZnO, A. Walsh, C. R. A. Catlow, M. Miskufova, A. A. Sokol, The Journal of Physics: Condensed Matter 23 **334217** (2011) in print

4. Stability of III-V nitride alloys (in preparation)

5. A QM/MM study of native defects in GaN (in preparation)

1 Introduction

The development of a commercial blue light emitting diode (LED; see Fig. 1) based on GaN by Nakamura [1] resulted in great interest in III-V nitrides, a term used to refer to compounds of a Group III element (usually Al, Ga or In in this context) and a group V element (N in this case). LEDs with a range of colour characteristics can now be manufactured by alloying III-V compounds in suitable proportions to generate the required band-gap for the emission of a particular colour (see Fig. 2). White light LEDs have the potential to replace conventional light bulbs and improve efficiency [2]. These compounds have found many other applications in semiconductor electronic and opto-electronic devices, which include laser diodes (LDs), as used in Blu-ray disc technologies for example, high electron mobility transistors (HEMTs), distributed Bragg reflectors (DBRs) and metal-oxide-semiconductor field-effect transistors (MOSFETs). More recently, there have been promising applications of GaN/InGaN in solar cells [3] and photodetectors [4].

Three methods have been widely used in the synthesis of GaN (and similarly AlN and InN): metal organic vapor phase epitaxy (MOVPE), which is the

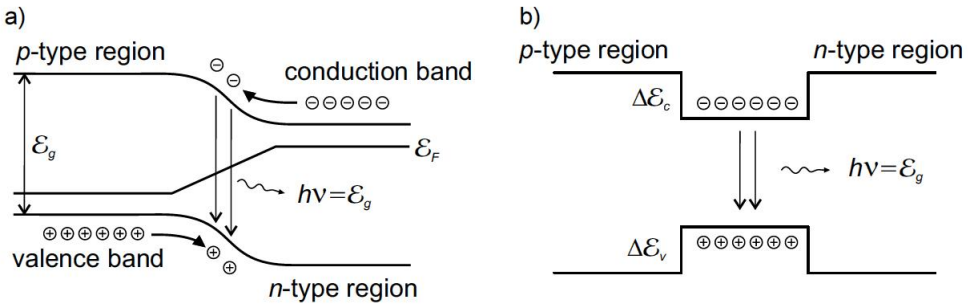


Figure 1: The operation of two types of LED. a) the p-n junction based LED. A forward bias is applied to the junction so that majority carriers (holes in the p-type material and electrons in n-type) cross the junction and become minority carriers on the other side, where they undergo radiative recombination and emit light with frequency $\nu = \epsilon_g/h$. ϵ_F is the Fermi level and ϵ_g is the band gap b) quantum well based LEDs. The quantum well confines electrons and holes and recombination occurs emitting light. From [5].

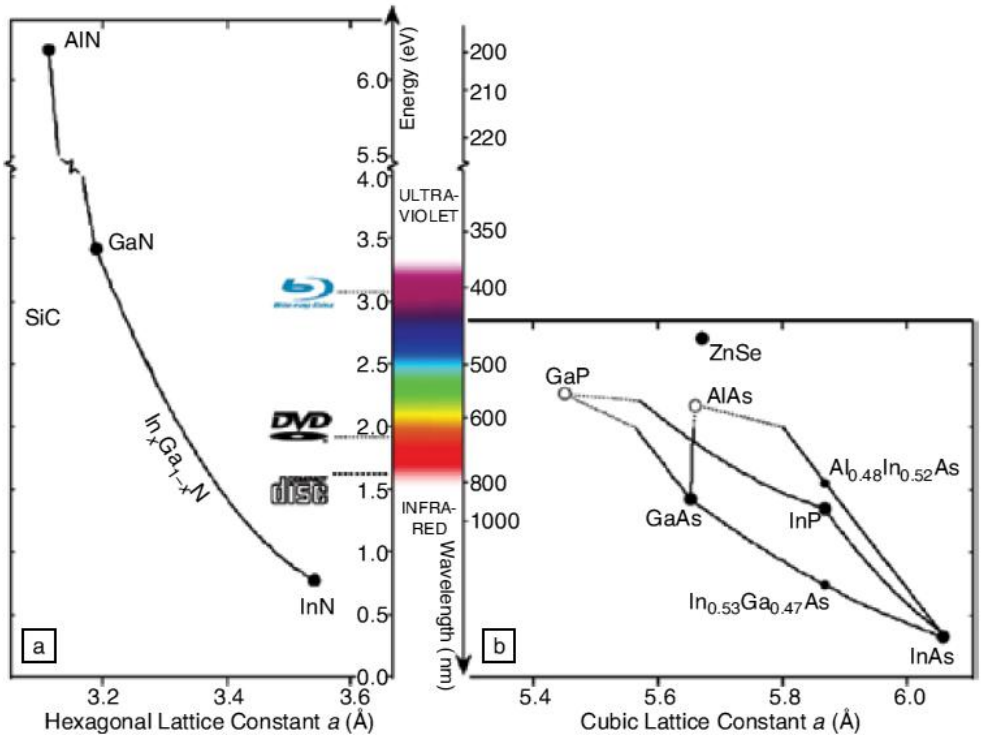


Figure 2: The band gaps vs. structural parameters of III-V wurtzitic nitrides on the left. The colour range of light emitted is shown in the middle along with the wavelengths used in some applications such as Blu-ray technology. On the right, the relationship between the band gap and structure parameters of some materials previously used in LEDs are shown. From [2].

variant of the metal organic chemical vapour deposition (MOCVD) used for epitaxial growth, hydride vapour phase epitaxy (HVPE) and molecular beam epitaxy (MBE). MOVPE and HVPE do not use high vacuum and the compound is grown by a reaction between injected gases. In MOCVD/MOVPE tri-methyl-Ga (or ocassionally tri-ethyl-Ga) reacts with ammonia and the product is deposited on wafers of substrate, such as sapphire, in a highly controlled environment. In HVPE, Ga reacts with HCl to give gallium chloride, which then combines with ammonia to create GaN. MBE uses a high vacuum and is based on physical deposition of atoms. The growth rates for this method are slower than the other two. A review of the growth methods can be found e.g. in [6] and [7].

GaN is most readily synthesised as an n-type material. Creating p-type GaN

has proven challenging and was finally synthesised by doping with Mg in 1989 [8].

Three phases of III-V nitrides are common. Wurtzite is the more stable phase under ambient conditions; zincblende is metastable and rocksalt is a high pressure phase. As wurtzite is the most stable phase under commonly encountered conditions, we will focus our attention on this polymorph in the present work.

In commercial applications and also experimental studies, the materials are usually present as thin layers on a substrate of another material. The properties of the material in such an epitaxial layer might be quite different from the bulk as strain can have a profound influence.

The electronic properties of the wide bandgap semiconductors are of particular importance in the aforementioned applications. The band structures have been widely studied and the work has yielded an understanding of important aspects of the behaviour of these materials. The electronic structure of the bulk materials is a useful basis for exploring their electronic properties. However, in addition to the bulk band structure, defects in the material, i.e. deviations from the perfect crystalline periodic structure, are important determinants of electronic behaviour. The theoretical prediction of the properties of GaN defects, such as their energetics, geometric configurations and electronic structure, is the main purpose of the present work.

To access the detailed electronic properties in a theoretical calculation, quantum mechanics has to be invoked. The famous Schrödinger equation forms the basis of quantum mechanics (QM). While the formulation is very elegant, the solution remains a major challenge for all but the simplest systems even with the advent of powerful computers to carry out the calculations at ever increasing speeds.

Hartree-Fock approaches (Section 2.2.1) suffer from accuracy problems due to their neglect of correlation effects and large computational cost due to the number of two electron integrals that have to be evaluated and have consequently not been widely used in modern studies of GaN.

One of the most successful approaches to making quantum mechanical calculations feasible for large assemblies of atoms, such as those in a periodic

solid, is density functional theory (DFT; Section 2.2.2). As the name suggests, it re-casts the Schrödinger equation in terms of electronic density, in theory vastly simplifying the calculations, although in practice this formalism brings its own set of challenges.

For crystalline materials such as III-V nitrides, we can utilise the underlying periodicity of the lattice in the DFT calculation to impose periodic boundary conditions (PBC; see Section 2.2.2) . The crystal is represented by an infinite array of identical repeating units ("supercells") with restrictions placed on the wavefunction at the boundary between the units to ensure physically correct behaviour. Such a treatment is suitable for the study of bulk properties but leads to difficulties when studying isolated defects or their complexes as the defects in neighbouring supercells will interact with each other unless the supercells are large enough for even the long-range electrostatic defect self-interactions between defects to be negligible. Furthermore, in the infinite periodic system, a charged defect in each supercell would lead to the system having infinite electrostatic energy. This problem was solved by Leslie and Gillan [9]. When the defect is not a classical point defect but has a dipole or higher order multipole terms associated with it, treating the electrostatic energy becomes much more difficult. Several methods have been introduced to deal with this problem, for example one due to Makov and Payne [10] but the problem has not been fully solved to date. A more detailed discussion of the problems arising from the electrostatic interactions in a periodic DFT calculation are discussed in [11]. A detailed analysis of the problems inherent in treating charged defects with periodic boundary conditions is given in [12], [13] and [14].

A further issue arises if the geometric deformations associated with the defect are large and the supercell size is small relative to them. The distortions will interact with each other at the boundary, affecting the result. More localised defects have larger deformations associated with them and require a better quality functional to describe them, usually hybrid functionals, which are very expensive for PBC DFT calculations. LDA (local density approximation) and GGA (generalised gradient approximation) functionals, commonly used in PBC DFT calculations, tend to underestimate the bandgap

significantly, potentially seriously distorting defect energy calculations. New hybrid functionals have been developed recently [15], which are computationally somewhat less expensive and their utilisation in studies of GaN has become feasible [16]. These functionals are better suited to the description of localised states than LDA and GGA functionals. Alternative methods, such as DFT+U, where U is an empirical parameter, have been also been proposed to deal with localised states in DFT while still using the LDA or GGA energy functionals. This approach, however, requires some *a priori* knowledge of the type of localisation required in a particular defect.

An alternative approach to defect calculations, adopted in the present work, was specifically developed for these purposes. A cluster including the defect and surrounding atoms is carved out and its electronic structure calculated. The remainder of the crystal is included at two levels of approximation. The atoms adjacent to the QM region are described with molecular mechanics (MM; Section 2.1) where specially parameterised functions, obtained by fitting to experimental data, describe the energy surface in the material. Atoms further away from the QM cluster interact with it via Coulomb forces. These far-away atoms are represented by a number of carefully calculated point charges chosen to reproduce the Madelung potential. This approach forms the basis of so-called QM/MM hybrid embedding methods, whose implementation forms the basis of Chapter 8.

In QM/MM methodology, atoms (ions) in the system interact via long range and short range forces. The latter are determined via correlation and exchange functionals in DFT and approximated with short range interatomic potentials in the MM portion of the calculation. In the case of charged species only, the long range interactions are the electrostatic Coulomb forces. Here, an important technical choice arises, viz. what charges to assign to the ions in a solid such as GaN.

The relative degree of ionicity and covalency can be difficult to ascertain [18]. Different measures of ionicity/covalency (e.g. difference in electronegativities, charge analysis [17]) give most compounds different mixtures of ionic/covalent character [18]. Based on these ionicity measures there is no consistent way of assigning ionic charge.

This distinction between ionic and covalent bonding is also important in deciding on the type of QM region termination in hybrid embedding methods. For covalent systems the QM region is terminated with so-called link atoms, which are usually capping hydrogens to terminate a severed covalent bond to atoms in the MM region. No such bonds exist in the case of ionic compounds, where we terminate the QM region with a pseudopotential, mimicking the effect of the electronic cloud surrounding an ion just outside the QM region. The pseudopotentials prevent unphysical spilling of electrons out of the QM cluster. QM region termination is further discussed in Section 2.3.1 of the present work.

GaN has a Pauling ionicity of 0.486 [20] and Phillips ionicity [21] of 0.500 [22] and yet another measure of ionicity, Majewski-Vogl ionicity is used as a basis for comparing GaN and AlN in [23]; [18] casts doubt on all these ionicity measures. An ionicity of 1 under both schemes implies 100% ionic character. Several MM models assign the ions in GaN non-formal charges, i.e. other than 3+ for gallium and 3- for N.[1] These models are often based on Mulliken analysis [17], e.g. [24] and [25] where charges 2+ (Ga) and 2− (N) are used although the Mulliken charges assigned to ions in different polymorphs of the same material can be quite different. For example, in wurtzite GaN the Mulliken charges have been calculated as +1.6 for Ga and -1.6 for N, whereas the corresponding values for the rocksalt phase are +2.5 and -2.5 [26], which presents difficulties when developing a potential to describe both the polymorphs. Further discussion of the issue of ionic charges in GaN can be found in [27], where a variable charge interatomic potential is developed. No clear choice of fixed partial charges for GaN is apparent. In the present work, we adopt a different approach and develop a formal charge model.

One technical issue, as mentioned above, is the choice of a QM region termination scheme. If we were to assign the partial charges 2+ and 2− to Ga and N, for example, based on the results of Mulliken analysis, we would implicitly assume a significant degree of covalency. In effect, a single electron would be assumed to be the basis of a bond to the four nearest neighbours. Such a choice would, however, present difficulties when dealing with atoms

[1] All the charges in this work are in the units of the electronic charge

on the boundary. The partial charges suggest that when a bond is cleaved a non-integer number of electrons (1/4 in the case of a 2+/2- charge state) is transferred. A link atom scheme would be necessary to saturate the dangling bonds. We believe that due to the significant ionic character of GaN, a link atom termination scheme, suitable for highly covalent systems, is not a good approach. Termination with pseudopotentials was therefore chosen.

We will treat the formal-charge assumption as one of our hypotheses. The results of our calculations will provide a test of this hypothesis. A failure of our fully ionic model correctly to account for a considerable covalent character of bonding would be expected to lead, among other potential problems, to high-energy electronic states on the boundary due to missing link-atom termination, which is necessary for covalent systems. In future work, a partial charge model using link atoms could be developed and comparisons made with our results to test our formal charge hypothesis further. Developing such a partial charge model is beyond the scope of this work.

A further reason for developing a formal charge model is the need to study alloys. The interatomic potential described in Chapter 5 was used to investigate ternary Al/Ga/In nitride alloys, which requires a common N-N interatomic potential, and therefore a common charge on the nitride ion. Since the three binary compounds have varying degrees of ionic character, and as has been mentioned before, even for a single material the measures differ in their assessment of ionicity, a formal charge model is a sound choice in this instance.

As a final remark on the issue of appropriate ionic charges, we note that it has been suggested that for a material with an appreciable degree of covalency, one can either use a partial charge model or a formal charge model with three-body terms (see Section 2.1) [24]. [28] developed a formal charge model for AlN but did not find it necessary to include three-body terms for this material.

In the present work, we focus our attention on GaN and discuss AlN and InN where appropriate. One of the main aims is to establish a methodology for modelling highly ionic solids and their alloys. In particular, we aim to validate a QM/MM method for defect calculations in such solids.

While QM/MM models do not have some significant problems that DFT calculations suffer from, such as the interaction of images in the neighbouring supercells and the band gap problem (provided that a hybrid functional is used in the QM/MM study), they are not without their limitations. Shallow defects do not lend themselves to study by QM/MM methods as the boundary conditions require that the electrons do not delocalise beyond the extent of the QM region. QM/MM methods can also be challenging to implement as a suitable interatomic potential is required. Distortions of the QM/MM boundary are observed if an incorrectly parameterised potential is used.
The remainder of this thesis is organised as follows.
Chapter 2 discusses the computational aspects of the work, which include QM methods (in particular DFT), MM methodology (including parameter fitting) and QM/MM hybrid schemes. General issues of geometry optimisation and energy minimisation are also discussed.
Chapter 3 gives greater detail on the theoretical background regarding various physical properties, including elastic constants and phonon frequencies. A brief summary of the theory of defects, cation ordering and phase separation is also given, the latter being relevant for the study of alloys.
Chapter 4 summarises the literature regarding methodological developments in interatomic potentials for the III-V nitrides and the QM/MM approach as well as defects in GaN. Experimental and computational studies of III-V nitride alloys are discussed, especially their stability with respect to phase separation and the evidence for cation ordering in the lattice.
We then turn to a description of a new formal charge interatomic potential model for GaN, AlN and InN in Chapter 5. The bulk properties calculated with the model are compared with other results as well as experimental data. These include elastic and piezoelectric constants, which are important as III-V nitrides are usually present in applications as thin films with considerable amounts of strain. The effect of this strain on the structural parameters and electric response to the strain are measured by the elastic and piezoelectric constants, respectively. Dielectric constants are another important property, which must be reproduced correctly if charged defects are to be studied. Phonon frequencies are a sensitive indicator of the quality of the

model. They form the microscopic basis for macroscopic thermodynamic properties, such as thermal expansion, which is particularly relevant in the high-operating-temperature applications of III-V nitrides as differential thermal expansion coefficients between substrate and epitaxial thin films result in large amounts of strain and concomitant engineering problems. Finally, the phase stability of the commonly found polymorphs is studied.
In Chapter 6, the model developed in the previous chapter is used to calculate the energies and structures of formal-charge native defects, interstitials and vacancies, in the three binary materials. The solution energies of the unlike cations in each compound are also calculated.
In Chapter 7, we turn our attention to ternary alloys, or solid solutions, of the III-V nitrides. Controlling the composition of an alloy allows one to engineer many of its properties, such as band gap, and alloys have found a number of industrial applications. In this chapter we consider their stability with respect to phase separation into their constituent components and investigate the energetics and structures of a number of ordered arrangements. As most practical uses of III-V nitrides involve highly strained epitaxial films, the effect of strain is also considered.
In Chapter 8, we take the potentials developed and thoroughly validated in the previous chapters and combine them with quantum mechanical techniques in a hybrid embedding scheme allowing us to calculate the energetics, geometries and electronic structure of defects in GaN. In addition to insights into the science of native defects in GaN, methodological issues regarding QM/MM implementation are discussed.
Finally in Chapter 9, the findings of this work are summarised and future directions of work indicated.

References

[1] p-GaN/n-InGaN/n-GaN double heterostructure blue-light-emitting diodes, S.Nakamura, M.Senoh, T.Muka, Jpn. J. Appl. Phys., Part 2, 32 **L8 -L11** (1993)

[2] Solid State Lighting, C. J. Humphreys, MRS Bulletin 33 **459 - 470** (2008)

[3] Design and characterization of GaN/InGaN solar cells, O.Jani, I.Ferguson, C.Honsberg, S.Kurtz, Appl.Phys.Lett. 91 **1 - 3** (2007)

[4] InGaN/GaN multi-quantum-well-based light-emitting photodetective dual-functional devices, C.Miao, H.Lu, D.Chen, R.Zhang, Y.Zheng, Front.Optoeletron.China 44 **441 - 443** (2009)

[5] Multiscale modelling of III-nitrides: from dislocations to the electronic structure, D.Holec, PhD thesis, University of Cambridge (2008)

[6] Growth and applications of Group III-nitrides, O Ambacher, J. Phys. D: Appl. Phys. 31 **2653 -2710** (1998)

[7] Fundamentals of semiconductors, P.Y.Yu, M.Cardona, Springer 2003

[8] P-Type Conduction in Mg-Doped GaN Treated with Low-Energy Electron Beam Irradiation (LEEBI), H.Amano, M.Kito, K.Hiramatsu, I.Akasaki, Jpn. J. Appl. Phys. 28 **L2112 -L2114** (1989)

[9] The energy and elastic dipole tensor of defects in ionic crystals calculated by the supercell method, M.Leslie, N.J.Gillan, J.Phys.C 18 **973 -982** (1985)

[10] Periodic boundary conditions in *ab initio* calculations, G.Makov, M.C.Payne, Phys.Rev.B 51 **4014 - 4022** (1995)

[11] Native defects and self-compensation in ZnSe, D.B. Laks, C.G. Vandewalle, G.F. Neumark, P.E. Blochl and S.T. Pantelides Phys. Rev. B 45 **10965 - 10978** (1992)

[12] Energy estimates for local chemical processes in condensed matter, C.Pisani, S.Casassa, Mol.Eng.7 **231 - 244** (1997)

[13] Issues in first-principles calculations for defects in semiconductors and oxides, R.M.Nieminen, Model.Simul.Mater.Sci.Eng. 17 **084011** (2009)

[14] Which electronic structure method for the study of defects: a commentary, W.R.L.Lambrecht, phys.stat.sol.B 1-12 **1547 - 1558** (2010)

[15] Screened hybrid density functionals applied to solids, J.Paier, M.Marsman, K.Hummer, G.Kresse, I.C.Gerber, J.G.Angyan, J.Chem.Phys. 124 **154709** (2006)

[16] Carbon impurities and the yellow luminescence in GaN, J. L. Lyons, A. Janotti, C. G. Van de Walle, APL 97, **152108** (2010)

[17] Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I, R.S.Mulliken, The Journal of Chemical Physics 23 **1833 - 1840** (1955)

[18] Ionicity in solids, C.R.A.Catlow, A.M.Stoneham, J.Phys.C:Solid state phys. 16 **4321 - 4338** (1983)

[19] Stability of a planar-defect structure of the wurtzite AlN $(10\bar{1}0)$ surface: Density functional study, H.Ye, G.Chen, Y.Wu, Y.Zhu, S.-H. Wei, PRB 80 **033301** (2009)

[20] Nature of the chemical bond, L.C.Pauling, 3rd Ed. (1960)

[21] Ionicity of the Chemical Bond in Crystals, J.C.Phillips, Rev.Mod.Phys.42 **317** (1970)

[22] Anisotropic thermal expansion in wurtzite type crystals, H.Iwanaga, A.Kunishige, S.Takeuchi, J.Mat.Sci. 35 **2451 - 2454** (2000)

[23] Physical properties of GaN and AlN under pressures up to 0.5 Mbar, P. Perlin, A. Polian, J.P. Itieb, I. Grzegorya, E. Litwin-Staszewskaa, T. Suskia, Physica B: Condensed Matter 185 **426 - 427** (1993)

[24] An interatomic potential study of the properties of gallium nitride, P.Zapol, R.Pandey, J.D.Gale, J.Phys.: Condens.Matter 9, **9517 - 9525** (1997)

[25] Classical simulations of the properties of group-III nitrides, J.A. Chisholm, D.W.Lewis, P.D.Bristowe, J.Phys.Condens.Matter 11 **L235 - L239** (1999)

[26] *Ab initio* study of high pressure phase transition in GaN, R.Pandey, J.E.Jaffe, N.M.Harrison, J.Phys.Chem.Solids 55 **1357-1361** (1994)

[27] Dynamic charge-transfer bond-order potential for gallium nitride, K. Albe, J.Nord, K.Nordlund, Philosophical Magazine 89 **3477 - 3497** (2009)

[28] Intrinsic disorder in aluminum nitride A.N,Cormack, J.Am.Ceram.Soc.72 **1730 - 1732** (1989)

2 Background on computational methods

In this chapter, some of the common methods for calculating the properties of materials are described. These can be divided into electronic structure techniques and approaches based on interatomic potentials. Various combinations of the methodologies have been described such as hybrid QM/MM methods, which will also be considered.

2.1 Molecular mechanics (MM) methods

In molecular mechanics methods, the potential energy of the interactions between different species is expressed via parameterised functions. The parameters and functional forms are chosen so that the physical properties of the system are well reproduced. These properties include the structural parameters which are determined at the equilibrium geometries, i.e. where the forces on all atoms are zero. The forces are determined as the negative derivatives of the energy with respect to the coordinates.
The functional form of the interactions depends on the system under consideration. For covalent systems, the potential energies are frequently expressed in terms of the bond bending and stretching. For highly ionic systems, such as III-V nitrides the interatomic potential is divided into a long range Coulombic part, describing the interaction of charged species, and a short range potential, which simulates the interactions other than Coulombic of the outer electrons of the ions.
The Coulomb interaction is given by

$$V_{\text{electrostatic}} = -\frac{q_1 q_2}{4\pi\epsilon_0 r} \tag{1}$$

where q_1 and q_2 are the ionic charges, ϵ_0 is the permittivity of free space and r is the interionic separation. The present potentials are formal charge models, i.e. all the cations are assigned charge 3+ and nitrogen ions are 3- as discussed in Chapter 1. Evaluating the total Coulomb energy for a periodic system is complicated as the number of interacting ions increases as r^3 with distance r

from a reference atom while the electrostatic potential decreases as $1/r$. The sum of electrostatic energy contributions is conditionally convergent. Various solutions have been proposed to this problem. The Ewald summation is the most commonly used [1], [2]; another method is due to Evjen [3].

A related concept to the electrostatic potential is the Madelung field. It can be determined for a single ion by summing up the electrostatic interactions with all other ions in the lattice, which are approximated as point charges. The sum of the interactions of the ion i with all the other ions in the lattice may be written as

$$V_i = \frac{e}{4\pi\epsilon_0} \sum_{j \neq i} \frac{z_j}{r_{ij}} \tag{2}$$

where z_j is the charge of the jth ion. Normalising to the nearest neighbour distance, r_0

$$V_i = \frac{e}{4\pi\epsilon_0 r_0} \sum_{j \neq i} \frac{z_j r_0}{r_{ij}} = \frac{e}{4\pi\epsilon_0 r_0} M_i \tag{3}$$

where

$$M_i = \sum_{j \neq i} \frac{z_j}{r_{ij}/r_0} \tag{4}$$

is the Madelung constant. The summation is challenging as the series is conditionally convergent as described above in the discussion of summation of the total electrostatic energy of the lattice, but again the Ewald technique may be used.

The above approximation treats the lattice ions as point charges. It is possible to generalise the expression and include multipole terms. Higher order Madelung constants are thus obtained [4], [5].

Next we turn to the short range interatomic potentials. Common forms include Buckingham, Born-Meyer, Morse and Lennard-Jones. In addition, the shell model simulates the interaction between the core and valence electrons of an ion via a spring.

The Buckingham potential is often used in modelling III-V nitrides. It has the form

$$V_{\text{Buckingham}} = Ae^{-r/\rho} - \frac{C}{r^6} \tag{5}$$

where r is the separation between the species, and A, ρ and C are parameters. The first term represents the repulsion between the electrons of neighbouring electron clouds, whilst the second term is equivalent to the r^{-6} term of the Lennard-Jones interaction (see below).

The Born-Meyer potential is a special form of the Buckingham, with $C = 0$.

A Morse potential may be used to describe more covalent-like interactions. In the fitting of the present potentials it was found that the N-N interaction is well-suited to this description.

$$V_{\text{Morse}} = D_e\left(\left(1 - e^{-a(r-r_e)}\right)^2 - 1\right) \tag{6}$$

This form was originally developed to describe diatomic molecules as it accounts for the harmonic oscillator behaviour of molecules as well as anharmonic effects. The D_e parameter is then the dissociation energy of the bond; r_e determines the position of the minimum of the function and a affects the width of the potential well.

Another commonly used function is the Lennard-Jones 12-6 potential.

$$V_{\text{Lennard-Jones}} = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} \tag{7}$$

where r is the interatomic distance and C_{12} and C_6 are empirically determined constants.

The C_{12} term represents repulsion due to the overlap of electronic clouds and is fundamentally due to the Pauli exclusion principle. There is no firm theoretical justification for the use of the 12 exponent but this model has been successfully used in a number of systems. The 12-term is also widely applied to avoid "Buckingham catastrophe" at the very short range, whereby the electrostatic attraction between oppositely charged species overcomes the exponential repulsion.

The r^{-6} has a solid theoretical foundation and is based on dispersion interactions. It is used for modelling the N-N interaction beyond the 1$^{\text{st}}$ neighbour shell in our model.

The above models are all two-body potentials, i.e. the interactions between

two ions are considered. Higher-order terms can also be included, which model the interactions between three ions and more. Three-body terms will include the angle subtended by the the vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{jk}$ where i, j and k are the three ions considered in the interaction. These terms model the directionality of covalent bonds in the material and have not been used in the present work although their inclusion might improve some aspects of the model and may be considered in future work.

The shell model [6] approximates the polarisation behaviour of electrons without using quantum mechanics and allows dielectric properties to be modelled. The polarisability α of an ion in the shell model is described by

$$\alpha = \frac{Y^2}{k} \tag{8}$$

The charge is split between the core and the massless shell of the ion, their sum adding up to the overall charge. The shell of charge Y is attached to the core with a spring constant k. The potential, V_2 due to the interaction is

$$V_2 = \frac{1}{2}kx^2 \tag{9}$$

where x is the displacement of the shell from the core. The N ion is a highly polarisable anion and needs this approach.

A quartic term in the shell displacement may be included to prevent excessive polarisation

$$V_4 = \frac{1}{4}k_4x^4 \tag{10}$$

The x^4 potential increases more rapidly than the x^2 term with increasing displacement from the core. This term effectively prevents the shells from moving too far off the core and hence dampens polarisation.

The MM software used in the present work is the General Utility Lattice Program (GULP), developed by Julian Gale [1].

2.1.1 Potential fitting

Like the rest of the MM work in this thesis, interatomic potential fitting was performed with the GULP software [1]. In this implementation, potential

parameters are fitted by minimising the sum of squares using a Newton-Raphson functional minimisation process (see, for example, [40] and a short summary below). The sum of squares, F, is defined as

$$F = \sum_{\text{all}k} w(f_{\text{calc, k}} - f_{\text{obs, k}})^2 \tag{11}$$

where w is the weighting factor, f_{calc} and f_{obs} are the calculated and observed quantities (e.g. structural parameters, elastic constants) respectively. Experimental properties provide indirect information about the energy surface. Near equilibrium, the first-order derivatives are zero by definition as $-E'(x)$ represents forces on the ions. The second order derivatives are related to elastic constants and phonon frequencies, which is why these experimental properties are a useful source of information. As an alternative, *ab initio* potential energy surfaces are sometimes used as "observables" although this approach was not used in the present work.
Relax fitting was used. In this form, the structure is optimised at every fitting step, in contrast to conventional fitting where observables are calculated for the observed structure which may be off the energy minimum.

2.1.2 Mott-Littleton method

The Mott-Littleton method is used to calculate the energies of point defects in MM as an alternative to the supercell approach (see Section 2.2.2). In the latter, a repeating supercell containing the defect is the basis of a periodic boundary condition calculation. For charged defects, this method introduces problems with interaction of defects with their images in neighbouring cells. The Mott-Littleton approach does not suffer from the problem arising from defect-defect interactions. In its original implementation [42], which did not employ computers, the equilibrium positions of the nearest neighbours of the defect centre (vacancy or interstitial) were calculated. The rest of the crystal was treated as a continuum whose polarisation contributes to the

defect energy a negative term first proposed by Jost [43]

$$E_{\text{Jost}} = -\frac{q^2}{2R}\left(1 - \frac{1}{\epsilon_0}\right) \tag{12}$$

The approach has evolved since then with advances in computational resources. The GULP implementation [1] of the methodology is discussed below.

Two concentric regions are defined, region 1 and 2, with region 2 further subdivided into 2a and 2b. The defect centre defines the centre of a sphere of radius r_1 containing region 1 and the defect usually lies in the defect centre. The region between radii r_2 and r_1 ($r_2 > r_1$) is region 2a. The rest of the crystal is defined as region 2b.

It is assumed that atoms in region 1 are strongly perturbed by the presence of the defect and fully relaxed during the calculation with respect to their coordinates. In region 2a, it is assumed that the perturbation is weak enough and that the response to the forces is harmonic. This approximation can only be employed when the structure is close enough to a minimum for the harmonic approximation to be valid.

Representing the coordinates of region 1 atoms by x and the displacements of region 2a atoms by ζ we obtain the following relationship for the total energy of regions 1 and 2a

$$E = E_1(x) + E_{12}(x, \zeta) + E_2(\zeta) \tag{13}$$

where E_1 and E_2 are the energies of regions 1 and 2, respectively, and E_{12} is the interaction energy between those two regions.

As mentioned before, the approximation in regions 2a stipulates that the energy of region 2 is a quadratic function of ζ

$$E_2(\zeta) = \frac{1}{2}\zeta^T W \zeta \tag{14}$$

At equilibrium, the first derivative with respect to region 2a displacements much be zero, i.e.

$$\frac{\partial E}{\partial \zeta} = 0 = \frac{\partial E_{12}(x,\zeta)}{\partial \zeta} + W\zeta \tag{15}$$

Combining Equations 13, 14 and 15 we obtain

$$E = E_1(x) + E_{12}(x,\zeta) - \frac{1}{2}\frac{\partial E_{12}(x,\zeta)}{\partial \zeta}\zeta \tag{16}$$

Thus the difficult E_2 term is eliminated. The total energy can then be minimised. There is a further technical issue with calculating E_{12}, viz. that the energy of the electrostatic interaction between regions 1 and 2 is performed using the Ewald summation, assuming a perfect lattice. Because ions in region 2a are no longer at their perfect lattice sites, a correction has to be applied.

The displacements of region 2a, ζ may be calculated by keeping in mind the quadratic approximation. Then

$$\zeta = -W^{-1}g \tag{17}$$

where W is the second derivative matrix and g is the matrix of forces on the ions. A commonly applied approximation equates g to the electrostatic force due to the defect species.

Region 2b interaction energies are calculated by implicitly considering the polarisation of the region due to the defect, which is based on the approach pioneered by Jost [43] although it does not treat the remainder of the crystal as a continuum but considers the polarisation of individual ions and performs a variant of Ewald summation.

Further discussion of the Mott-Littleton approach can be found, for example, in [7] and [8].

2.2 QM methods

While MM methods allow one to model material behaviour at the atomic level, electronic properties can be only crudely approximated, e.g. by the

shell model. In order to understand the electronic structure, quantum mechanics has to be invoked. Different *ab initio* methods have been developed and are briefly summarised below. Further details can be found in standard texts, e.g. [9].
In quantum mechanics, the Schrödinger equation is solved. In its time-independent form it can be written as

$$\hat{\mathbf{H}}\Psi = E\Psi \tag{18}$$

where $\hat{\mathbf{H}}$ is the Hamiltonian and Ψ is the wavefunction. E, the eigenvalue in this equation, represents the energy of the system.
In the absence of external fields, the electronic Hamiltonian operator is, in atomic units

$$\hat{\mathbf{H}} = -\frac{1}{2}\sum_{i}^{\text{electrons}} \nabla_i^2 - \sum_{i}^{\text{electrons}}\sum_{A}^{\text{nuclei}} \frac{Z_A}{r_{iA}} + \sum_{i\neq j}^{\text{electrons}} \frac{1}{r_{ij}} + \sum_{A\neq B}^{\text{nuclei}} \frac{Z_A Z_B}{r_{AB}} \tag{19}$$

A number of formalisms and approximations have developed to solve this equation for many electron systems, such as tight-binding and Green's functions methods. However, most modern studies of III-V nitrides have focussed on Hartree-Fock and Density Functional Theory (DFT) derived approaches.

2.2.1 Hartree Fock methods

The wavefunction Ψ determines the energy E via

$$E = \frac{\langle\Psi|\mathbf{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{20}$$

The wavefunction has to be antisymmetric with respect to the exchange of two electronic coordinates, a requirement imposed by the Pauli exclusion principle. This can be satisfied by expressing the wavefunction in terms of

Slater determinants S_D

$$S_D = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \dots & \phi_N(r_1) \\ \phi_1(r_2) & \phi_2(r_2) & \dots & \phi_N(r_2) \\ \dots & \dots & \dots & \dots \\ \phi_1(r_N) & \phi_2(r_N) & \dots & \phi_N(r_N) \end{vmatrix} \tag{21}$$

where $\phi_n(r_m)$ is the n^{th} spinorbital, i.e. the product of electron spatial orbitals and spin function, which is an appropriate approximation in the non-relativistic limit, and r_m is the radius-vector of electron m. The spin orbitals are orthonormal $\langle\phi_i|\phi_j\rangle = \delta_{ij}$. In the Hartree-Fock approach, an approximation of a single Slater determinant is made.
It can be shown that this constrained optimisation problem (keeping the one electron orbitals orthonormal) leads to the Hartree-Fock equations

$$\mathbf{F(r_1)}\phi_\mathbf{i}\mathbf{(r_1)} = \epsilon_\mathbf{i}\phi_\mathbf{i}\mathbf{(r_1)} \tag{22}$$

where ϕ_i' are the canonical molecular orbitals and F_i is the Fock operator.

$$\mathbf{F(r_1)} = \mathbf{h(r_1)} + \sum_\mathbf{j}^\mathbf{N}(\mathbf{J_j(r_1)} - \mathbf{K_j(r_1)}) \tag{23}$$

The first term has its origins in the description of the kinetic energy of electron i in the field of the nuclei of the system,a,

$$\mathbf{h(r_i)} = -\frac{\mathbf{1}}{\mathbf{2}}\nabla_\mathbf{i}^\mathbf{2} - \sum_\mathbf{a}\frac{\mathbf{Z_a}}{\mathbf{|R_a - r_i|}} \tag{24}$$

J is the Coulomb term, a classical repulsion between like charges

$$\mathbf{J_i(r_1)}|\phi_\mathbf{j}\mathbf{(r_1)}\rangle = \langle\phi_\mathbf{i}\mathbf{(r_2)}|\ \frac{\mathbf{1}}{\mathbf{|r_1 - r_2|}}\ |\phi_\mathbf{i}\mathbf{(r_2)}\rangle|\phi_\mathbf{j}\mathbf{(r_1)}\rangle \tag{25}$$

$\mathbf{K}$ is the exchange term, with no classical analogy. It arises as a result of the antisymmetry requirement of the wavefunction

$$\mathbf{K_i}(\mathbf{r_1})|\phi_\mathbf{j}(\mathbf{r_1})\rangle = \langle\phi_\mathbf{i}(\mathbf{r_2})|\,\frac{\mathbf{1}}{|\mathbf{r_1}-\mathbf{r_2}|}\,|\phi_\mathbf{j}(\mathbf{r_2})\rangle|\phi_\mathbf{i}(\mathbf{r_1})\rangle \tag{26}$$

Since the Fock operator depends on the final wavefunction the equations have to be solved in a self-consistent manner to obtain a final wavefunction and energy.

Various approximations have been implemented, including neglecting some overlap integrals and parameterising others. These adaptations form the basis of semiempirical methods and result in increased computational speed but possible loss of accuracy.

Other methods derived from the Hartree Fock approach, such as the configuration interaction, try to correct for the neglect of electron correlation. They are beyond the scope of this work and are discussed in more detail in [9].

2.2.2 Density functional theory(DFT)

DFT removes the problem of dealing with the complicated many-body interaction among electrons by replacing it with a single-body calculation where the electronic density n is the key variable. While a wavefunction in HF-based methods depends on 3N coordinates (4N including spin) where N is the number of electrons, density only depends on three coordinates, independent of the number of electrons, in theory vastly reducing the computational effort.

A proof by Hohenberg and Kohn [10] underlies the method, stating that the ground state electronic energy is fully determined by the density. The problematic part in DFT involves mapping the density onto the energy, i.e. determining the energy functional. The second important theorem [10], based on the more general variational principle, states that the true ground state electron density is that which gives the lowest possible energy, which is equal to the ground state energy.

In order for these theorems to be of practical use in reducing the computa-

tional effort in electronic structure calculations, the Kohn-Sham equations reformulate the problem in terms of N non-interacting one-electron systems. The interactions of an electron with all the other electrons, rather than being included in the wavefunction via its dependence on all the electronic coordinates, are now dealt with via the effective potential term in the Hamiltonian. The problem is then reformulated as the Kohn-Sham equations [11]

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\boldsymbol{r}) \right] \psi_i = \epsilon_i \psi_i \tag{27}$$

where

$$v_{eff}(\boldsymbol{r}) = v(\boldsymbol{r}) + \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r}' + v_{XC}(\boldsymbol{r}) \tag{28}$$

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2 \tag{29}$$

$v(\mathbf{r})$ is the potential due to nuclei and any external fields and the second term in Eq. 28 is the classical Coulomb interaction. This reformulation puts the two difficult aspects into the v_{XC} term. Firstly, it is difference between the kinetic energy of the non-interacting electron gas, described by the KS equations, and the true kinetic energy of the fully interacting system. Secondly, all non-classical aspects of the electron-electron interaction are included in this term.

The exchange term arises naturally in HF when employing a Slater determinant formulation as a result of applying the antisymmetrising operator. However, even including this term does not yield the correct energy of the system, which is lower than predicted by HF. The difference between the true minimum and the HF minimum is termed the electron correlation energy. Electrons are correlated and as a result they are further apart than predicted by the HF wavefunction. The correlation between electrons in the same molecular orbital is larger than inter-orbital correlation. Since electrons with the same spin are not allowed in the same orbital due to the Pauli exclusion principle, the correlation between opposite spins is larger than same spin correlation, which has no intra-orbital contribution. Opposite spin cor-

relation is termed Coulomb correlation and same spin correlation is known as Fermi correlation. Due to correlation effects there is a reduced probability of finding another electron in the immediate vicinity of one, i.e. there is hole around it.

The exact form of the exchange correlation functional is unknown and the approximations developed to describe it are summarised later in this section. The effective potential is a function of density, but density is also calculated via equation 29, which in turn depends on the effective potential. The equations are solved in the self-consistent field (SCF) approach. The majority of computational approaches to solving the KS equations are iterative. The initial guess KS orbitals are integrated to obtain the density function, which in turn is used to determine v_1. Kohn-Sham equations are then solved to yield a new set of orbitals (with new expansion coefficients), which then determine the new density and v_2. If the difference between v_1 and v_2 is too large, the process is repeated until self-consistency is achieved.

The Kohn-Sham (KS) orbitals can be either described numerically or as an expansion over basis functions (plane waves, Gaussians, etc.).

The total energy of the system can be written as

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{30}$$

where T_S is the non-interacting kinetic energy, E_{ne} is the energy of the interaction between nuclei and electrons, E_{xc} is the exchange and correlation energy and J is the electron Coulomb term.

The boundary conditions in DFT calculations of crystalline solids can be divided into two classes. Under periodic boundary conditions a repeating supercell unit imposes conditions on the boundaries between supercell. As an alternative is a cluster approach without any periodicity, a method that is extensively used in this thesis.

Functionals

A number of functionals have been developed to describe the exchange and correlation interaction. Some of the most common ones used in studies of GaN are local density, gradient corrected and hybrid functionals.

In the local density approximation (LDA), the electron density is assumed to be slowly varying in space and the exchange energy can be written as

$$E_{xc}^{LDA}[\rho] = \int \epsilon_{xc}(\rho_\uparrow, \rho_\downarrow)\rho(\mathbf{r})\mathbf{dr} \tag{31}$$

where ϵ_{xc} is the exchange-correlation energy density. Reasonably accurate formulae for ϵ_{xc} have been determined from Monte Carlo simulations. VWN is an example of LDA functional.

In the generalised gradient approximation (GGA) the energy density is a function of the first spatial derivative of density a well as the density itself.

$$E_{xc}^{GGA}[\rho] = \int \epsilon_{XC}(\rho_\uparrow, \rho_\downarrow, \nabla\rho_\uparrow, \nabla\rho_\downarrow)\rho(\mathbf{r})\mathbf{dr} \tag{32}$$

PW91 and LYP are examples.

Meta-GGA methods employ a second derivative of the density in the expansion and the Kohn-Sham kinetic energy density (or its Laplacian). Asymptotically corrected methods aim to reproduce correctly the $1/r$ asymptotic behaviour of the exchange energy density. Hybrid functionals use exact exchange energy, calculated from HF methods, to improve accuracy in addition to any number of explicit energy functionals. B3LYP is an example.

The functional relationships between the energy density and the electron density and its derivatives are often quite complicated and depend on the exact functional employed.

Basis sets

Basis sets provide a way of expressing a wavefunction ψ in terms of known functions χ_i, i.e. the basis set weighted by variable coefficients c_i

$$\psi = \sum_i c_i\chi_i \tag{33}$$

This expansion is fully accurate if the basis set is complete. In computational practice it is impossible to have a complete basis set, which would require an infinite number of basis functions. Both the actual number and quality of the basis functions determine the accuracy of the calculation.

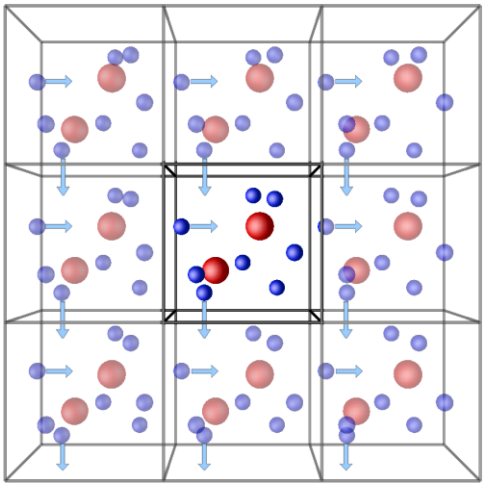


Figure 3: Periodic boundary conditions. The central box represents a repeating cell, which is not necessarily the unit cell of the material, and has to be chosen carefully so that the underlying inherent periodicity of the lattice is not disturbed. A supercell with a number of unit cells is often chosen, e.g. a $2 \times 2 \times 2$ will contain two unit cells along each lattice vector. The arrows represent forces on the particles. From [12].

The most commonly used basis sets are Slater-type orbitals (STOs), Gaussian-type orbitals (GTOs) and plane wave basis sets.

Plane waves are used with periodic boundary conditions (Fig. 3), which employ an infinitely repeating supercell. The wavefunction Ψ can be expanded due to Bloch's theorem as

$$\Psi_i(r) = e^{ik \cdot r} f_i(r) \tag{34}$$

where k is the electronic momentum and

$$f_n(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}} \tag{35}$$

Here G are reciprocal lattice vectors. Periodic boundary conditions and plane waves were not used in the present work but are common in other studies of GaN defects.

STOs have the form

$$\chi^{STO}_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r} \tag{36}$$

where N is a normalisation constant, $Y_{l,m}$ are spherical harmonics and ζ is the orbital exponent.
A GTO can be written as

$$\chi^{GTO}_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2} \tag{37}$$

In Cartesian coordinates

$$\chi^{GTO}_{\zeta,l_x,l_y,l_z}(x,y,z) = Nx^{l_x}y^{l_y}z^{l_z}e^{-\zeta r^2} \tag{38}$$

The type of orbital is determined by $l_x + l_y + l_z$ where l are the orbital quantum numbers. For a p-type orbital the sum is equal to 1, for example. There are five d-type components of a GTO in polar coordinates ($Y_{2,2}$, $Y_{2,1}$, $Y_{2,0}$, $Y_{2,-1}$ and $Y_{2,-2}$) whilst there are six Cartesian components (x^2, y^2, z^2, xy, xz and yz). It can be shown that these six functions may be transformed into five spherical d-functions and one s-function ($x^2 + y^2 + z^2$). Similarly, the 10 Cartesian f-functions can be shown to transform into seven spherical f-functions and one set of three p-functions. In computational methods the excess functions are often removed, speeding up the calculation.
STOs are in theory more accurate than GTOs as they are better behaved near the nucleus and they do not fall off as quickly as the GTOs hence the tails of the functions are better represented. However, it is more efficient to calculate two-electron integrals with GTOs and they are often preferred in computational chemistry.
The minimal basis set employs the same number of functions as there are electrons in the neutral atom. Doubling the minimal basis set results in a double zeta (DZ) type basis set. If only the functions for the valence electrons are doubled a split valence basis is formed, sometimes referred to as valence double zeta. Triple and higher order zeta basis sets have also been developed.

Basis functions with higher angular momenta than those occupied in the ground state of the atom are also included to describe polarisation effects (e.g. p functions for hydrogen, d functions for carbon). Inclusion of too many polarisation functions with a small number of sp-functions might however result in artefacts as the higher angular momentum functions compensate for inadequacies in the sp basis.

Another common addition to a basis set are diffuse basis functions. These have small ζ parameters in the exponential, resulting in a slower decay and a better description of the tail of the wavefunction. This may be necessary for modelling loosely bound electrons, e.g. in anions, or if the property being studied depends on the wavefunction tail, e.g. polarisability.

To improve the computational efficiency of the basis set, it is often contracted. The contraction is usually applied to the chemically largely inert but energetically important core electrons. It is performed by fixing some of the c_i coefficients so that they are no longer subject to calculation via the variational principle. Three basis functions χ_1, χ_2 and χ_3 with three variable coefficients c_1, c_2 and c_3 may be contracted to one basis function $c_{\text{contr}}(A_1\chi_1 + A_2\chi_2 + A_3\chi_3)$ where A_i are constants and c_{contr} is a variable coefficient. The basis functions in the contraction are called primitive functions.

Pseudopotentials

In atoms with a large number of core electrons the basis functions representing the core electrons are both expensive and unnecessary and they are often replaced with a pseudopotential, also known as an effective core potential (ECP) (see [13], [14], [15], [16], [17], [18] and [19]). In deriving preudopotentials, an accurate all-electron wavefunction is calculated. The valence orbitals are then replaced with pseudorbitals which are equivalent to the true wavefunction outside the core region (at distances $> R_C$ outside the nucleus) but lack the nodal structure in the core region (see Fig.4). A pseudopotential is then introduced which results in Schrödinger equation yielding the pseudo wavefunctions and the correct Kohn-Sham eigenvalues. The pseudopotential is usually expanded in terms of analytical functions such as Gaussians.

A good pseudopotential needs to be transferable, i.e. applicable in a range of

atomic/molecular environments different from those it was originally fitted in.

A variety of approaches to constructing ECPs have been developed. In general, pseudopotentials are non-local, i.e. the radial form is a function of the angular momentum l. Kleinman and Bylander separated the local and non-local parts of the pseudopotential operator [20]. Hay and Wadt proposed a semilocal scheme [19]. Other models were developed by Huzinaga [21], Phillips and Kleinman [16] and Durand and Barthelat [22]. Norm-conserving pseudopotentials such as [14] require that the integrated charge within R_C matches that of a full-electron calculation.

In addition to reducing the computational expense of QM calculations, pseudopotentials are used to create a boundary between QM and MM subsystems in QM/MM methods (see below).

Pseudopotentials, or effective core potentials, can be divided into small core

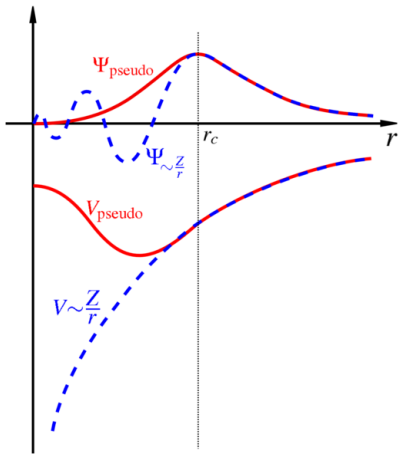


Figure 4: Comparing a wavefunction in a potential with a wavefunction in a pseudopotential.r_c is the cut-off radius. From [23]

potentials describing only the deepest core electrons and large core potentials which describe all the electrons bar the valence ones.

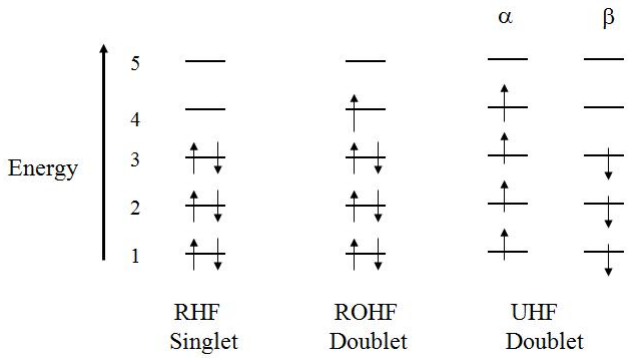


Figure 5: RHF, ROHF and UHF treatment of spin. From [24].

2.2.3 Spin in QM computational methods

Two electrons may be present in each orbital. The resulting spin-orbital is written as a product of the spatial wave function and the spin function α or β. If an even number of electrons are present and they have a singlet-type wavefunction (a closed shell system), i.e. each electron is paired up in an orbital with another electron, the spatial parts of the wavefunction of the two electrons in the same orbital are commonly restricted to be identical, which is known as Restrictred Hartree-Fock (RHF). Open shell-systems may also be restricted in such a way, resulting in Restricted Open-Shell Hartree-Fock (ROHF). More commonly for open shell systems though, the two spatial wavefunctions in the same orbital are not restricted, leading to Unrestricted Hartree Fock (UHF). The differences between the different configurations and the associated energy levels are illustrated in Fig. 5. The same concepts apply in DFT calculations.

In open shell systems, spin density is defined as the total density of electrons with spin α less the total density of electrons with spin β.

2.3 QM/MM methods

QM/MM methods aim to overcome the insufficient accuracy of MM and computational expense of QM methods by combining the two to create a potentially powerful and yet fast method. The system being studied is parti-

tioned between a chemically active region treated at the QM level and system remainder studied by MM. Further details on the partitioning of the system are given in Chapter 8.

2.3.1 Termination of the QM region

The methodologically difficult part is the treatment of the interface between the two subsystems, which can be achieved either by introducing extra nuclear centres (link atoms) or a boundary region.
Link atom schemes are well suited to the description of covalent systems, especially organic molecules (see e.g [25], [26], [27]). Variations on the method such as scaled position link atom method (SPLAM) [28], adjusted connection atoms [29] and double link atoms [30]) have been proposed.
Boundary region methods usually employ pseudopotentials, or pseudobonds in the case of covalent systems [31]. The Hamiltonian can then be rewritten as

$$H = \left(-\frac{1}{2}\sum_{i}^{N_{eff}} \nabla_i^2 \right) + \left(\sum_{i\neq j}^{N_{eff}} \frac{1}{r_{ij}} - \sum_{i}^{N_{eff}} \sum_{\alpha\in QM} \frac{Z_\alpha}{r_{\alpha i}} - \sum_{i}^{N_{eff}} \sum_{\beta\in MM} \frac{q_\beta}{r_{\beta i}} \right) + \\ + \left(\sum_{i}^{N_{eff}} \sum_{\gamma\in Y_{ps}} V_{Y_{ps}}(r_{i\gamma}) \right) + \left(\sum_{\alpha_1\neq\alpha_2\in QM} \frac{Z_{\alpha_1} Z_{\alpha_2}}{r_{\alpha_1\alpha_2}} + \sum_{\alpha\in QM,\beta\in MM} \frac{Z_\alpha q_\beta}{r_{\alpha\beta}} \right) \quad (39)$$

Here the first bracket represents the kinetic energy of N_{eff} electrons. These are QM region electrons, and, depending on the implementation, the valence electrons of the boundary atoms Y_{ps}. The interactions between the boundary atoms may be treated at the QM level, in which case they enter into the Hamiltonian as additional α atoms, or at the MM level, corresponding to β description in Eq. 39. The second bracket is the Coulomb interaction between the electrons in the system and of the electrons with the nuclear charges α in the QM region as well as the charges β of the MM region. The third term describes the interaction of the electrons with the ECPs $V_{Y_{ps}}$ centred on atoms Y_{ps}. The final term is a constant that accounts for the Coulomb

interaction of the nuclear charges in the QM region with themselves as well as the charges in the MM region.
Variations on the pseudopotential theme have been introduced. The quantum capping potential (QCP) [32] reduces the number of valence electrons (from four to one in the case of carbon) and introduces extra terms into the potential to account for this modification.

2.3.2 Energy expression

There are two main approaches to calculating the total energy of a QM/MM system: additive and subtractive schemes.
The total energy in an additive scheme can be written as

$$E_{total} = E_{MM}(X) + E_{QM}(Y + B) + E_{QM-MM}(Y + B, X) - E_{\text{corr}} \qquad (40)$$

where E_{MM} is the MM energy of the atom set X, E_{QM} is the QM energy of the atom set Y and the boundary (or link) atoms B and E_{QM-MM} is the energy of interaction between the QM and MM atoms. Corrections to the energy, E_{corr} are sometimes applied, for example a Jost correction (see Section 8.2.2).
In a subtractive scheme the energy may be written as

$$E_{\text{total}} = E_{\text{MM}}(X + Y + B) + E_{\text{QM}}(Y + B) - E_{\text{MM}}(Y + B) \qquad (41)$$

An example a subtractive scheme is ONIOM [33] , [34], [35], [36], [37], including approaches with more than two layers.
Additive schemes only will be used in the present work.

2.3.3 Interaction between QM and MM subsystems

The interaction between the QM and MM regions in an ionic solid take the form of the usual Coulomb and short range terms, although the parameters of the short range potential may have to be adapted relative to a pure MM model.

A further consideration is the treatment of polarisation in the QM/MM model. Here three different models have been developed: mechanical, electrostatic and polarised embedding.
In mechanical embedding, the electrostatic interaction between the QM and MM regions is confined to the Coulomb interaction between the ionic charges and is treated at the MM level of approximation, which introduces a constant term in the Hamiltonian. The electrons in the QM region are not polarised by the MM region, a serious drawback in the treatment of highly ionic systems. Furthermore, the method requires an accurate set of parameters, which might be difficult if the charge distribution in the QM region is altered from the bulk charge distribution, as is usually the case in the course of a chemical reaction.
Electrostatic embedding improves on the shortcomings of the previous approach by incorporating the interaction between the electrons and MM charges in the one-electron terms in the Hamiltonian. As a result the electrons in the QM region can respond to changes in the ionic charge distribution in the MM region. The MM region electrostatic contribution is no longer a constant additive term in the total energy expression. The MM region is, however, not polarised by the QM fragment.
The next logical step is to include the MM polarisation in the formalism. This is known as polarised embedding. The ionic shells of the shell model, as described above, respond to electronic polarisation in the QM region. The equilibrium polarisation is found by solving for the QM and MM region polarisations self-consistently. This methodology was used in the current work.
When link atoms are used with electrostatic and polarised embedding charge modifications on the QM-MM boundary might be required, including selective deletion of atomic charges and charge shifting. [38] discusses these issues further.
In addition to the long-range Coulomb interactions, short-range interactions between the QM and MM regions are included at the MM level of approximation. A more detailed description of our QM/MM implementation and the interactions between the different regions is decribed in Chapter 8.

2.4 Geometry optimisation

A number of methods are available for energy minimisation. The system at the energy minimum is said to have an optimised geometry, so energy minimisation and geometry optimisation are equivalent terms. The same principles apply to geometry optimisation in MM and QM/MM codes.
It has been found that methods involving the gradients of the function being optimised are efficient at finding the minimum of a 3D function. The minimum can be found by using the first order derivatives of a function; following the negative gradient leads to the solution. This method is known as "steepest descent". More sophisticated methods usually converge in fewer iterations but each step is computationally more demanding. Two popular variants, Newton-Raphson (and related approaches) and conjugate gradients are briefly summarised below. Standard texts such as [40] discuss optimisation techniques in more detail.
The potential energy can be regarded as a Taylor expansion around the equilibrium geometry.

$$E(\mathbf{x} + d\mathbf{x}) = E(\mathbf{x}) + E'(\mathbf{x})d\mathbf{x} + \frac{1}{2}d\mathbf{x}^T E(\mathbf{x})''d\mathbf{x} + \dots \tag{42}$$

where the $E'(\mathbf{x}) = \mathbf{g}$ and $E''(\mathbf{x}) = \mathbf{H}$ matrices contain first and second-derivatives respectively. In the Newton-Raphson algorithm the vector in the direction from the current geometry towards the energy minimum is approximated as

$$d\mathbf{x} = -\mathbf{H}^{-1}\mathbf{g} \tag{43}$$

The inversion of the Hessian, $\mathbf{H}$, is computationally demanding. A number of techniques have been developed to avoid calculating it at every optimisation step. These methods update the inverse Hessian matrix approximately based on the value of the coordinates and $\mathbf{g}$, the gradient vector. Examples of schemes for updating the inverse Hessian are Davidon-Fletcher-Powell (DFP) and Broydon-Fletcher-Goldfarb-Shanno (BFGS), the latter being the default in GULP [1], the software used in the current work. The exact Hessian is recalculated when the software deems the approximate one no longer appro-

priate or after a specific number of cycles. This approach only works when the starting geometry is close to a minimum.
The Hessian might be very expensive to calculate at all. In that case, it can be set to a unit matrix initially and subsequently updated using an update scheme such as BFGS. Alternatively, a method which does not require a Hessian and uses only first derivatives, such as conjugate gradients (CG, see below) can be employed.
When starting from a point sufficiently far from the minimum, Eq.43 becomes increasingly approximate. A further danger is that a maximum rather than a minimum is found. To improve on the method, a line search is performed in the direction of the search vector $d\mathbf{x}$

$$d\mathbf{x} = -\alpha \mathbf{H}^{-1}\mathbf{g} \tag{44}$$

α is chosen so that the energy along the direction dx is minimised. Once the minimum along that search vector is found, a new search vector is calculated.
An alternative to the Newton-Raphson scheme is the CG method. It is a generalisation of the steepest descent method, which starts with a direction that has the largest gradient $\mathbf{d_0} = -\nabla \mathbf{f}$. The following steps are orthogonal to the previous one, i.e. $\mathbf{d_i} \cdot \mathbf{d_j} = \delta_{\mathbf{ij}}$. CG also starts off with the steepest descent direction $\mathbf{d_0}$. Once the minimum has been found, the algorithm chooses another direction to search along using information from the previous steps to increase efficiency. The step direction in cycle i is found via

$$\mathbf{d_i} = -\mathbf{g_i} + \beta_i \mathbf{d_{i-1}} \tag{45}$$

The choice of β varies in different methods, such as Fletcher-Reeves, Polak-Ribiere and Hetenes-Schiefel. Full details can be found, for example, in [41].
Methods using second derivatives, such as Newton-Raphson are particularly suitable for small systems due to their efficiency. As the system size increases, calculating and inverting the Hessian becomes increasingly expensive and the memory requirements will rise as well. For very large systems, either the unit Hessian method, or ultimately the conjugate gradients approach, have to be

used. Furthermore, Newton-Raphson methods become less useful when the structure being optimised is outside the harmonic region [1].

References

[1] General Utility Lattice Program, J.D.Gale `https://www.ivec.org/gulp/help/gulp3.4_manual.pdf`

[2] Die Berechnung optischer und elektrostatischer Gitterpotentiale, P. P. Ewald, Ann. Phys. 64 **253 - 258** (1921)

[3] On the Stability of Certain Heteropolar Crystals, H. M. Evjen, Phys. Rev. 39 **675 - 687** (1932)

[4] The equivalent charge concept and its application to the electrostatic energy of charges and multipoles, E. F. Bertaut, J. Phys. (Paris) 39 **1331 - 1348** (1978)

[5] Methods of Calculating the Crystalline Electric Field, J. Kanamori, T. Moriya, K. Motizuki, T. Nagamiya J. Phys. Soc. Jap. 10, **93 - 102** (1955)

[6] Theory of the Dielectric Constants of Alkali Halide Crystals, B. G. Dick, Jr, A. W. Overhauser , Phys. Rev. 112 **90 - 103** (1958)

[7] A comparison of defect energies in MgO using Mott-Littleton and quantum mechanical procedures, R.W.Grimes, C.R.A.Catlow, A.M.Stoneham, J.Phys.Condens.Matt. 1 **7367 - 7384** (1989)

[8] The Mott-Littleton method:an introductory survey, A.B.Lidiard, J.Chem.Soc.Faraday.Trans.89 **341 - 349** (1989)

[9] Introduction to Computational Chemistry, F.Jensen, (1999)

[10] Inhomogeneous electron gas P, Hohenberg, W.Kohn. Physical Review 136 **B864 - B871** (1964)

[11] Self-Consistent Equations Including Exchange and Correlation Effects, W.Kohn, L.J.Sham. Physical Review 140 **A1133 - A1138** (1965)

[12] `http://isaacs.sourceforge.net/phys/pbc.html`

[13] *Ab initio* pseudopotentials for electronic structure calculations of polyatomic systems using density functional theory, M.Fuchs, M.Scheffler, Comp.Phys.Comm. 199 **67 - 98** (1999)

[14] Generalized norm-conserving pseudopotentials, D.R.Hamann, Phys.Rev.B 40 **2980 - 2987** (1989)

[15] New method for calculating wave functions in crystals, C.Herring, Phys.Rev. 57 **1169 - 1177** (1940)

[16] New method for calculating wave function in crystals and molecules, J.C.Phillips, L.Kleinman, Phys.Rev.116 **287 - 294** (1959)

[17] Efficient pseudopotentials for plane-wave calculations, N.Troullier, J.S.Martins, Phys.Rev.B 43 **1993 -2006** (1991)

[18] Pseudopotentials that work: From H to Pu, G.B.Bachelet, D.R.Hamann, M.Schlüter, Phys.Rev.B 26 **4199 - 4228** (1982)

[19] *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg, P.J.Hay, W.R.Wadt, J.Chem.Phys.82 **270 - 283**(1985)

[20] Efficacious form for model pseudopotentials, L.Kleinman, D.L.Bylander, Phys.Rev.Lett.48 **1425 - 1428** (1982)

[21] Atomic and molecular calculations with the model potential method.I, V. Bonifacic, S.Huzinaga, J.Chem.Phys.60 **2779 - 2786** (1974)

[22] New atomic pseudopotentials for electronic structure calculations of molecules and solids, P.Durand, J.C.Barthelat, Chem.Phys.Lett.27 **191 - 194** (1974)

[23] `http://en.wikipedia.org/wiki/Pseudopotential`

[24] Introduction to Computational Chemistry, NSF Computational Nanotechnology and Molecular Engineering Pan-American Advanced Studies Institutes (PASI) Workshop January 5-16, 2004, NSF Computational Nanotechnology and Molecular Engineering Pan-American Advanced Studies Institutes (PASI) Workshop January 5-16, 2004, A.S. Ichimura

[25] Combining *ab initio* techniques with analytical potential functions for structure predictions of large systems: method and application to crystalline silica polymorphs , U.Eichler, C.M.Kölmel, J.Sauer, J.COmp.Chem.18, **463 - 477** 1996

[26] A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulation, J.Comp.Chem.11, **700 - 733** (1990)

[27] A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl^+Cl^-$ exchange reaction and gas-phase protonation of polyethers, U.C.Singh, P.A.Kollman, J.Comp.Chem 7(6), **718 - 730** (1986)

[28] A hybrid method for solutes in complex solvents:Density functional theory combined with empirical force fields, M.Eichinger, P.Tavan, J.Hutter, M.Parrinello, J.Chem.Phys. 110, **10452 - 10467** (1999)

[29] Adjusted connection atom for combined quantum mechanical and molecular mechanical methods, I.Antes, W. Thiel, J.Phys.Chem. 103 **9290 - 9295** (1999)

[30] Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method, D.Das, K.P.Eurenius, E.M.Billings, P.Sherwood, D.C.Chatfield, M.Hodoscek, B.R.Brooks, J.Chem.Phys.117 **10534 - 10547** (2002)

[31] A pseudobond approach to combining quantum mechanical and molecular mechanical methods, Y.Zhang, T.Lee, W.Yang, J.Chem.Phys. 110 **46 - 54** (1999)

[32] Simple one-electron quantum capping potentials for use in hybrid QM/MM studies of biological molecules, G.A.DiLabio, M.H.Hurley, P.A.Christiansen, J.Chem.Phys.116 **9578 - 9584** (2002)

[33] ONIOM: a multilayered integrated MO + MM method for geometry optimisations and single point energy predictions. A test for Diels Alder reaction and Pt(P (t-Bu)$_3$)$_2$ + H_2 oxidative reaction, M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, J. Phys. Chem. 100, **19357 - 19363** (1996)

[34] IMOMM: a new integrated ab initio + molecular mechanics geometry optimisation scheme of equilibrium structures and transition states, F. Maseras, K. Morokuma, J. Comp. Chem.16, **1170 - 1179** (1995)

[35] Implementation of the IMOMM technology for performing combined QM/MM molecular dynamics simulations and frequency calculations, T. K. Woo, L. Cavallo, T. Ziegler, Theor. Chem. Acc. 100, **307 - 313** (1998)

[36] The IMOMM method opens the way for the accurate calculation of real transition metal complexes, F. Maseras, Chem. Commun. **1821 - 1827** (2000)

[37] The IMOMO method: the integration of different levels of molecular orbital approximations for geometry optimisation of large systems: test for n-butane conformation and S_N2 reaction: RCl+Cl-, S. Humbel, S. Siebel, K. Morokuma, J. Chem. Phys. 105,**1959 - 1967** (1996)

[38] QM/MM: what have we learned, where are we, and where do we go from here?, H.Lin, D.G.Truhlar, Theor.Chem.Acc 117 **185 - 199** (2007)

[39] GULP: A computer program for the symmetry-adapted simulation of solids, J.D.Gale, J.Chem.Soc.,Faraday Trans. 93 **629 - 637** (1997)

[40] Chapter 14, Introduction to Computational Chemistry, F.Jensen, John Wiley & Sons (1999)

[41] Numerical methods for scientists and engineers, Volume 1, H. M. Antia, Birkhäuser

[42] Conduction in polar crystals, I.Electrolytic conduction in solid salts, N.F.Mott, M.J.Littleton, Trans.Far.Soc.34 **485 - 499** (1938)

[43] Diffusion and electrolytic conduction in crystals (ionic semiconductors) W.Jost, J.Chem.Phys.1 **466 - 475** (1933)

3 Introduction to material properties

In this chapter, the relevant physical properties of III-V nitrides are reviewed, including bulk properties such as structure and phase stability. We also consider defect properties and phase separation in alloys.

3.1 Structural properties

III-V nitrides crystallise in three polymorphs: the hexagonal wurtzite structure (Fig. 9, space group P63/mc) and two cubic phases, zincblende (Fig. 9, $F\bar{4}3m$) and rocksalt (space group $Fm\bar{3}m$). The structural parameters of the unit cells and the fractional coordinates of the ions within them are tabulated in Table 1. There are four atoms in the unit cell of wurtzite and two in the cubic polymorphs. Wurtzite has three material-dependent structural parameters (a, c and u) and the other two polymorphs have only one.

The wurtzite and zincblende structures are both based on tetrahedra. The difference between these two polytypes lies in the rotation of successive tetrahedra along the c-direction (see Fig. 6). Wurtzite has ABABAB... type stacking. In zincblende the third layer is rotated by 60° in the ab-plane with

Table 1: Unit cell and the fractional coordinates of the three common GaN polymorphs

Structural parameters	Angles	Fractional coordinates
Wurtzite		
$a = b \neq c$	$\alpha = \beta = 90°$, $\gamma = 120°$	Ga 2/3, 1/3, 1/2 Ga 1/3, 2/3, 0 N 1/3, 2/3, u N 2/3, 1/3, 1/2+u
Zincblende		
$a = b = c$	$\alpha = \beta = \gamma = 90°$	Ga 0, 0, 0 N 1/4, 1/4, 1/4
Rocksalt		
$a = b = c$	$\alpha = \beta = \gamma = 90°$	Ga 0, 0, 0 N 1/2, 1/2, 1/2

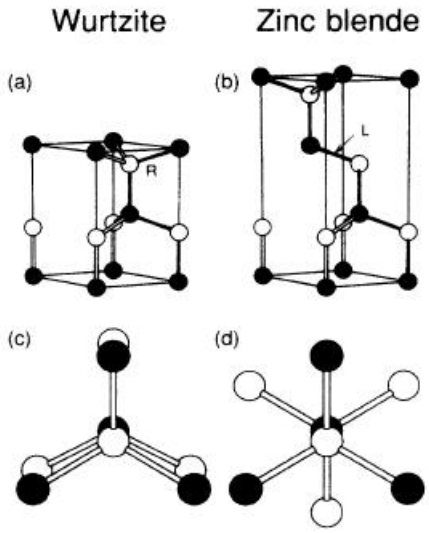


Figure 6: A diagram showing the difference between the wurtzite and zincblende polymorphs. a) and b) indicate the handedness of the successive layers, right (R) handed for wurtzite and left (L) handed for zincblende. The resulting views along the z-direction are shown in c) and d). From [1].

respect to wurtzite resulting in ABCABCABC... stacking. Other types of stacking are possible as shown in Fig. 8. This type of polytypism is very common in silicon carbide, where a wide range of polytypes is known.
The stacking differences result in wurtzite having hexagonal packing and zincblende being cubic (see Fig 9). Rocksalt is also cubic as can be seen in Fig. 7, but has the higher coordination number of 6.

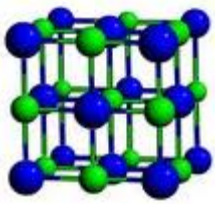


Figure 7: Rocksalt structure from [2]. The blue atoms are cations and the green atoms are nitrogens.

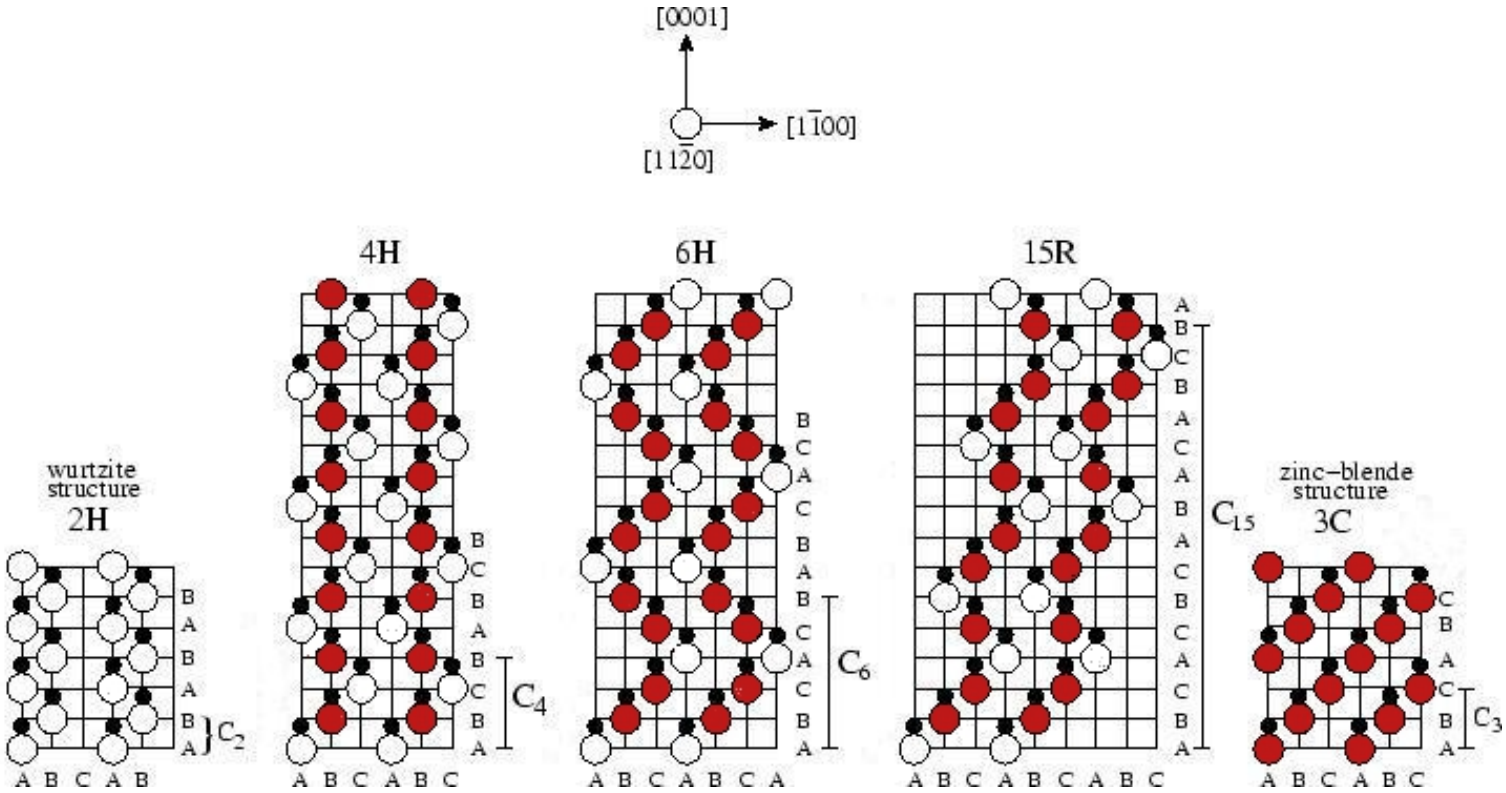
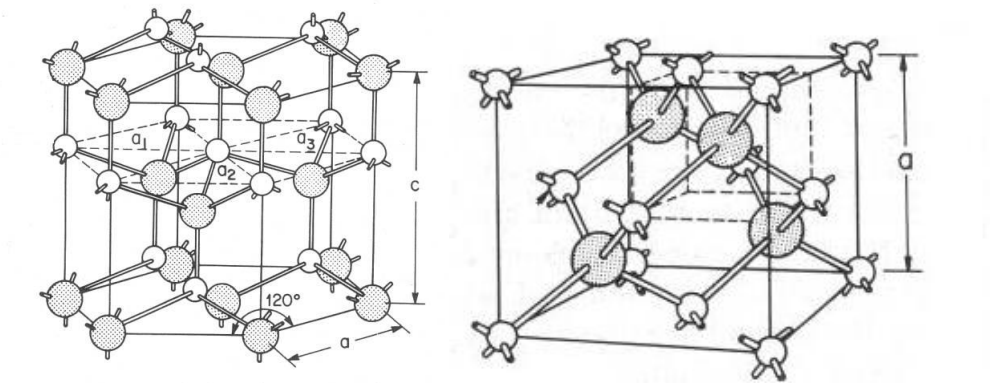


Figure 8: A diagram showing some possible III-V compound polytypes based on tetrahedra. White circles represent R-handed (wurtzite-like) tetrahedra and red circles show L-handed (zincblende-type) tetrahedra. Based on [3].



(a) wurtzite (b) zincblende

Figure 9: Two common polymorphs of III-V nitrides. The shaded spheres are cations and the white spheres are nitrogen ions. The hexagonal packing of wurtzite and cubic packing of zincblende is shown. From [4]

When indexing the Miller planes in wurtzite and zincblende a rotation is applied to the zincblende unit cell with respect to the one presented in Fig. 6. This rotation is shown diagrammatically in Fig. 10. The layer of like cations (or analogously, anions) has the crystallographic index (0001) in wurtzite, which after the rotation becomes (111) in zincblende. The (111) plane in zincblende is sometimes referred to as the pyramidal plane.

3.2 Elastic properties

Understanding the elastic properties is essential for the applications of materials such as the III-V nitrides in electroluminescent devices, where they are present as epitaxial thin films. The mismatches in the lattice constants between successive layers result in significant strain, especially at high temperatures, due to differential thermal expansion coefficients of the film and the substrate. The elasticity of the material determines the response to this strain.

The elastic constant tensor c_{ijkl} is defined as

$$\sigma_{ij} = c_{ijkl}\epsilon_{kl} \tag{46}$$

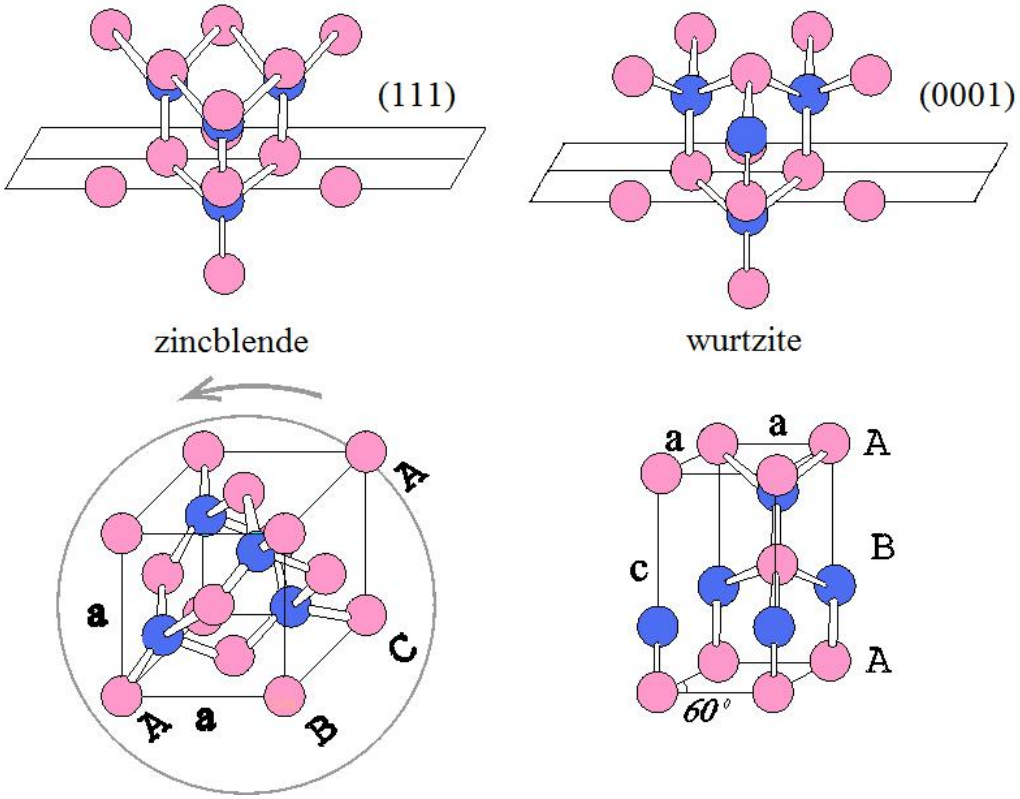


Figure 10: The difference between zincblende and wurtzite and their crystallographic planes. Ga in pink and N in blue. Note that the zincblende unit cell is rotated with respect to wurtzite so that the (111) zincblende plane is equivalent to the (0001) plane in wurtzite. From [5].

where σ_{ij} is the stress and ϵ_{ij} is the strain. As there are three coordinates, there are in principle $3\times3\times3\times3 = 81$ components of c_{ijkl}. Due to symmetry some components are equal and the elastic constant tensor is reduced to a 6×6 matrix. However, depending on the symmetry of the crystal some of the elements of the matrix are equal and the number of independent constants is further reduced. In wurtzite only the C_{11}, C_{12}, C_{13} and C_{44} are independent. A cubic crystal (e.g. zincblende) has only three independent elastic constants C_{11}, C_{12} and C_{44}.

The elastic constant matrix can be related to the internal energy U of the lattice via

$$C_{ij} = \frac{1}{V}\left(\frac{\partial^2 U}{\partial\epsilon_i \partial\epsilon_j}\right) \tag{47}$$

Strain ϵ is defined as

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i}\right) \tag{48}$$

where u_i is the deformation in r_i direction, and r_i (i = 1,2,3) are the coordinates. Equation 47 is used in MM calculations to calculate elastic constants. In the present work we model the behaviour of defectless bulk materials. Experiments are often performed on polycrystalline samples or thin films with significant concentrations of defects. Single crystals of III-V nitrides are difficult to grow. Furthermore, much of recent research has focussed on developing epitaxial films rather than bulk nitrides. These factors might explain some of the discrepancy between our model and experiment.

3.3 Dielectric properties

The response of a crystal to an electric field is in many ways analogous to its elastic reponse. Elastic constants measure the response to stress; dielectric constants measure how the polarisation of the material changes with applied electric field. Just like the elastic response, dielectric properties are frequency dependent. The intrinsic (high-frequency) behaviour corresponds to no relaxation involving ionic displacements whereas relaxation occurs in response to a low frequency electric field.

The electric field experienced by an ion in a crystal is a sum of any applied electric field and the field due to the displacement of the ions and electrons in the crystal. The relationship between this so-called local field, $\mathbf{E^{loc}}$ and the applied field may be expressed as

$$\mathbf{E^{loc}}(\mathbf{r}) = \frac{\epsilon + 2}{3}\mathbf{E}(\mathbf{r}) \tag{49}$$

where ϵ is the dielectric constant of the medium. The static and high frequency extremes of the dielectric function are usually quoted. In the high frequency case, only the electrons are able to respond to the electric field as the electric field oscillates faster than the maximum ionic phonon. In the static response, both the electrons and ions respond to the electric field. The

quality of the calculated dielectric constants is a sensitive indicator of the reliability of the model as second derivative matrices are used in the calculation.

The relationship between ϵ and the polarisability α is given by the Clausius-Mossotti relation

$$\frac{\epsilon(\omega) - 1}{\epsilon(\omega) + 2} = \frac{4\pi\alpha}{3v} \tag{50}$$

where v is the volume of the primitive unit cell. The polarisability is a sum of the polarisability due to the displacement of positive ions, α^+, that of negative ions, α^- and electronic polarisability

$$\alpha = (\alpha^+ + \alpha^-) + \frac{e^2}{M(\bar{\omega}^2 - \omega^2)} \tag{51}$$

Here M is the ionic mass and $\bar{\omega}$ is a vibrational frequency characteristic of the crystal. α can be seen as a link between the theoretical polarisability of the shell model (Section 2.1) and the macroscopic high frequency dielectric constant.

It can be shown that the Lydanne-Sachs-Teller relationship holds between longitudinal and transverse modes (ω_L and ω_T respectively)

$$\omega_L^2 = \frac{\epsilon_0}{\epsilon_\infty}\omega_T^2 \tag{52}$$

where ϵ_0 is the static dielectric constant and ϵ_∞ is the high frequency dielectric constant. This relationship is often useful in developing potentials as fitting the dielectric constants is often easier than matching phonon frequencies.

Whilst the dielectric properties describe the response of any material to an applied electric field, some materials show spontaneous polarisation (in the absence of an external field) due to the non-zero electric dipole moment associated with their unit cell, which is known as *ferroelectricity*. Wurtzite structures exhibit spontaneous polarisation while the cubic polymorphs do not.

3.4 Piezoelectric properties

Piezoelectricity refers to the ability of a material to generate an electric field in response to applied strain. The reverse piezoelectric effect refers to the generation of strain when an electric field is applied.
The relationships between the tensors of strain S, stress T, the electric charge density displacement D and electric field E can be expressed via piezoelectric coefficients

$$S = s^E T + d^t E \tag{53}$$

$$D = dT + \epsilon^T E \tag{54}$$

where d is the tensor of the direct piezoelectric effect and d^t is the tensor of the converse piezoelectric effect. The T and E superscripts indicate constant stress and electric field respectively.

3.5 Phonons

The lattice dynamics of a crystal are closely linked to its thermodynamic quantities. Atoms in the unit cell oscillate with respect to each other in different modes given by the symmetry of the cell. The normal modes of oscillation of the wurtzite cell are shown in Fig. 11. The oscillations in the xy-plane are doublets because the a and b directions are equivalent, thus giving two degenerate modes.

The lowest three energy phonons are the acoustic branches corresponding to the transmission of sound waves. The other modes, six in number for wurtzite, are termed optical modes. The highest frequency modes are those in which the neighbours move in the opposite direction to each other. In lower frequency modes, neighbours oscillate in phase. Zincblende has two atoms per unit cell hence six phonon branches (three acoustic and three optical).
Oscillations along the c-direction are the A and B modes. The modes with ions oscillating in the xy-plane are the E modes. The number of unique

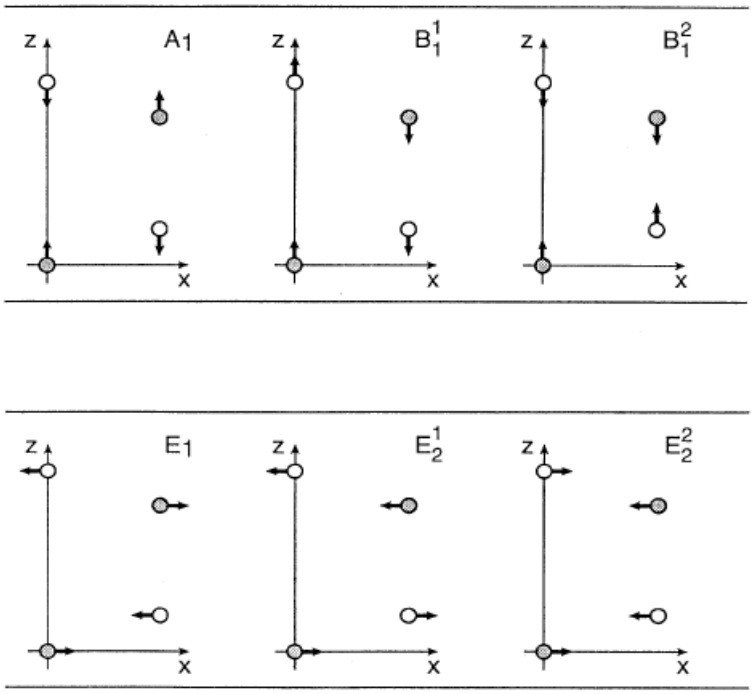


Figure 11: The normal modes of oscillation of a wurtzite lattice. The grey circles are cations and the white circles represent N ions. The arrows indicate the direction of relative motion in each mode. From [16]

optical modes is reduced from nine to six as the ab-plane modes are doubly degenerate due to the equivalence of the a and b directions. However, the degeneracy is broken for some of these modes for certain wavevectors $\mathbf{k}$[2] due to LO/TO (longitudinal/transverse) splitting. A LO/TO clasification is assigned to modes according to whether the atomic displacements are perpendicular or parallel to the direction of the wavevector. For example, approaching the Γ point[3] from the z-direction will give an A_1 longitudinal mode and two E_1 transverse modes with the same frequency (a doublet). Approaching the Γ point along a wavevector in the xy-plane will give an A_1 transverse mode and one E_1 transverse and one longitudinal mode.
The phase relationship between the oscillations in neighbouring unit cells is

[2]The wavevector determines the phase relationship between oscillations in neighbouring cells.

[3]The Γ point has the wavevector $\mathbf{k} = (\mathbf{000})$.

given by the wavevector **k**. The dependence of phonon frequencies ω on **k** is the dispersion relation of the phonons. The phonon frequencies are most commonly experimentally measured at the Γ point where all the unit cells oscillate in sync, i.e. $\mathbf{k} = (\mathbf{000})$. The Brillouin zones, i.e. the unit cells in reciprocal space, and their important symmetry points, of which the Γ point is one, for wurtzite and zincblende are shown in Fig. 12.

The density of states is a convenient concept in the computation of ther-

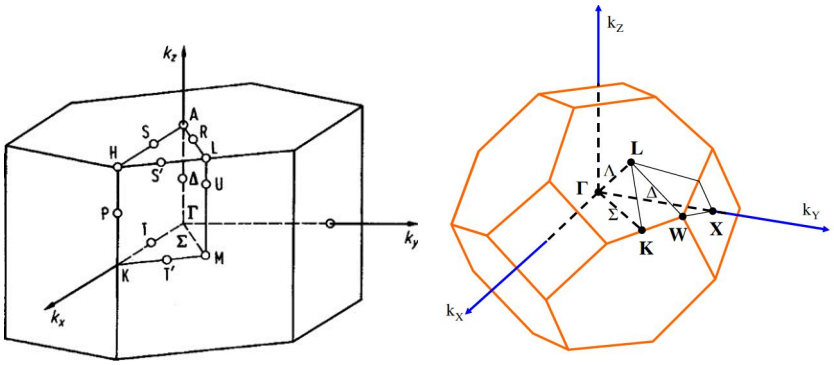


Figure 12: The Brillouin zone for wurtzite (left) [8] and zincblende (right) [9] with indicated special directions:
wurtzite K $(1/3, 1/3, 0)$, M $(1/2, 0, 0)$, A $(0, 0, 1/2)$, H $(1/3, 1/3, 0.5)$, L $(0.5, 0, 0.5)$
zincblende X $(1, 0, 0)$, L $(1, 1, 1)$, K $(3/4, 3/4, 0)$, W $(1, 1/2, 0)$

modynamic quantities that involve integration over k-space, such as specific heats. This integral may then be expressed as an integral over frequencies rather than wavevectors. The density of states gives the weighting factor in the integration for any frequency ω, which may be understood as the number of available states in the interval ω to $\omega + \delta\omega$.

In computational implementations, numerical summing scheme over points in a Monkhorst-Pack mesh is usually carried out [10] to obtain the density of states. In the limit of the grid spacing going to zero, this converges on the true result.

3.6 Thermodynamic properties

The quantum theory of the harmonic crystal predicts a temperature dependence of the specific heat c_v, given by

$$c_v = \frac{1}{V} \sum_{\mathbf{k},\mathbf{s}} \frac{\partial}{\partial T} \frac{\hbar\omega_s(\mathbf{k})}{e^{\beta\hbar\omega_s(\mathbf{k})} - 1} \tag{55}$$

where V is the volume, T is the temperature, $\omega_s(\mathbf{k})$ is the frequency of mode s at wavevector $\mathbf{k}$. In the low temperature limit this can be approximated as a T^3 dependence in insulators and $AT + BT^3$ in metals, tailing off at high temperatures to a constant. At intermediate temperatures the Debye and Einstein models are used to describe specific heats.

Whilst many material properties can be explained by the harmonic approximation of the energy expression, the description of certain properties dictates the inclusion of cubic and quartic terms; higher order terms are generally neglected. For example, the volume dependence of phonon modes is a direct consequence of the anharmonicity of the ionic interaction energy. Other phenomena such as thermal expansion and the non-identity of constant volume and constant-pressure specific heats can only be explained by anharmonic terms in the expansion.

It can be shown (e.g. Chapter 25 in [11]) that the thermal volume expansion coefficient α, may be written as

$$\alpha = \frac{\gamma c_v}{3B_0} \tag{56}$$

where c_v is the specific heat at constant volume, B_0 is the bulk modulus and γ is the overall Grüneisen parameter. γ can be written in terms of the Grüneisen parameters of normal mode $\mathbf{k}$,s as

$$\gamma = \frac{\sum_{\mathbf{k},\mathbf{s}} \gamma_{\mathbf{k},\mathbf{s}} c_{vs}(\mathbf{k})}{\sum_{\mathbf{k},\mathbf{s}} c_{vs}(\mathbf{k})} \tag{57}$$

The Grüneisen parameter of a mode is defined by

$$\gamma_{\mathbf{k,s}} = -\frac{V}{\omega_s(\mathbf{k})}\frac{\partial\omega_s(\mathbf{k})}{\partial V} = \frac{B_0}{\omega_{\mathbf{k,s}}(P=0)}\frac{\partial\omega_{\mathbf{k,s}}}{\partial P} \tag{58}$$

where $\omega_s(P=0)$ are the phonon frequencies at zero pressure.
The specific heat at constant volume of the mode is

$$c_{vs} = \sum_{\mathbf{k}} \frac{\hbar\omega_s(\mathbf{k})}{V}\frac{\partial}{\partial T}n_s(\mathbf{k}) \tag{59}$$

where $n_s(\mathbf{k}) = [\mathbf{e}^{\beta\omega_\mathbf{s}(\mathbf{k})} - \mathbf{1}]^{-\mathbf{1}}$.
As the main contribution of the temperature dependence of the thermal expansion coefficient comes from the c_v term, it can be expected that α will show similar temperature behaviour, i.e.

$$\alpha \propto T^3, T \to 0; \tag{60}$$

$$\alpha = \text{constant}, T \gg \Theta_D \tag{61}$$

3.7 Band structure

The computational methods employed to calculate the band structure are discussed in detail in [7]. The calculated band structure of GaN is shown in Fig. 13.
Electrons at zero temperature reside in the orbitals which altogether form the valence band. The conduction band comprises the unoccupied orbitals, which can nonetheless be partially filled by excitation and result in electronic conductivity, hence the name. The highest energy point of the valence band is known as the valence band maximum (VBM) and the bottom of the conduction band is the conduction band minimum (CBM). The band gaps form the basis of important electrical properties of the material. Wurtzite GaN has a bandgap of 3.505eV [13] and AlN 6.2eV[14]. The band gap of InN [15] has recently been determined to be around 0.7eV, although a wide range of values between 0.65 and 2.3eV has been reported over the last decades [16].

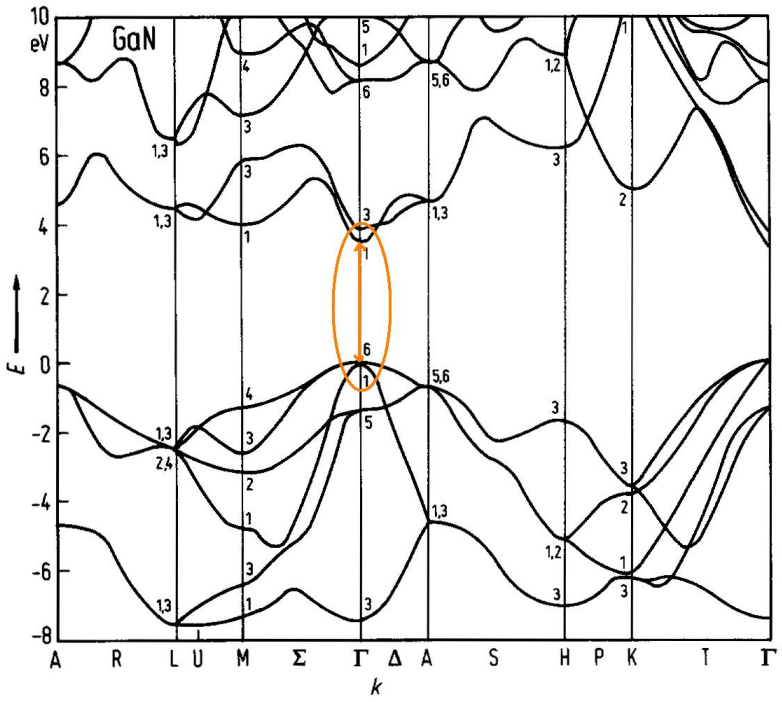


Figure 13: The band structure of GaN. Calculated in [12] using an empirical pseudopotential method. The orange arrow indicates the direct band gap of the material.

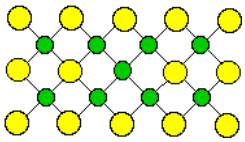


Figure 14: Antisite. The central anion (yellow) is replaced by a cation (green)

3.8 Defects

3.8.1 Formation energy

While a perfect crystal is a useful approximation for the study of many material properties, real crystals contain significant numbers of imperfections. These defects can be surface, line or point defects. In this section, only point defects will be considered.

There are three main types of defect - vacancies (absence of ions in a lattice), interstitials (presence of extra ions) and antisites (see Fig. 14). It can be

shown [2] that the concentration of a defect at thermodynamic equilibrium, $c = n/N$, where n is the number of defects and N is the number of lattice ions, is given by

$$c = N_{\text{sites}} N_{\text{config}} e^{-(E_f/kT)} \tag{62}$$

Here N_{sites} is the number of possible defect sites per unit volume, N_{config} is the number of equivalent configurations. $N_{\text{config}} = 1$ if no symmetry breaking occurs. E^f is the energy required to create the defect and T is the temperature. This equation in effect states that the higher the formation energy the lower the concentration of the defect.

The mobility of the defects has to be large enough to allow their concentration to equilibrate. Kinetic barriers, for example, might keep the concentrations of high energy defects above the equilibrium value. If the system is too far off thermodynamic equilibrium, formation energies lose their usefulness in predicting defect concentrations.

The formation energy depends on how the defect was generated and the relative chemical potentials of the species involved. In the case of charged species, the formation energy will also be a function of the Fermi level, E_F. The formation energy of a defect can be written, in the Zhang-Northrup formalism [17], as

$$E^f(q, E_F) = E^{\text{tot}}(q) - E^{\text{tot}}_{\text{bulk}} - \sum_i \Delta n_i \mu_i + q E_F \tag{63}$$

where $E^{tot}(q)$ is the total energy of a system with a defect in charge state q , E^{tot}_{bulk} is the total energy of the bulk, Δn_i is the change in the number of atoms of type i during the formation of the defect, μ_i is their chemical potential and E_F, the Fermi level, is the electron chemical potential with respect to the valence band maximum. $\Delta n_i > 0$ if an atom i is being added < 0 if it is being removed. The chemical potential represents the energy of the reservoir of the relevant species, which acts as a source or a sink.

For example, the formation energy of an oxygen substitutional atom with

charge q in GaN lattice, removing a nitrogen atom, can be written as

$$E^f(GaN:O_N^q) = E_{tot}(GaN:O_N^q) - E_{\text{bulk}}^{\text{tot}} - \mu_O + \mu_N + qE_F \tag{64}$$

where $E_{tot}(GaN:O_N^q)$ is a quantity that may be taken from a defect calculation such as a QM/MM calculation. The defectless bulk energy E_{bulk}^{tot} is similarly calculated. The chemical potentials will depend on the growth conditions.

At equilibrium, the chemical potentials of Ga and N are related to the chemical potential of GaN via

$$\mu_{\text{GaN}} = \mu_{\text{Ga}} + \mu_{\text{N}} = E_{\text{tot}}[\text{GaN}] \tag{65}$$

The chemical potentials have to satisfy the condition $\mu_{\text{N}} \leq \mu_{\text{N[N}_2\text{]}}$ and $\mu_{\text{Ga}} \leq \mu_{\text{Ga[bulk]}}$. Here $\mu_{\text{N[N}_2\text{]}}$ is the energy of N in the N_2 molecule and $\mu_{\text{Ga[bulk]}}$ is the chemical potential bulk Ga metal. If these boundary conditions are not satisfied the system would decompose into Ga and N. The chemical potential of GaN is given by the total energy of a two-atom unit of bulk GaN, $E_{\text{tot}}[\text{GaN}]$.

In the nitrogen-rich limit, the chemical potential of the nitrogen atom is half the binding energy of a nitrogen molecule. The chemical potential of Ga is then determined in the nitrogen-rich limit from $E_{\text{tot}}[\text{GaN}]$ and the N chemical potential via Eq.65 as follows. $E_{\text{tot}}[\text{GaN}]$ can be expressed as

$$E_{\text{tot}}[\text{GaN}] = \mu_{\text{Ga[bulk]}} + \mu_{\text{N[N}_2\text{]}} + \Delta H_f[\text{GaN}] \tag{66}$$

where $\Delta H_f[\text{GaN}]$ is the enthalpy of formation of GaN. Combining Eq. 65 and 66 and eliminating $E_{\text{tot}}[\text{GaN}]$ gives

$$\mu_{\text{Ga}} + \mu_{\text{N}} = \mu_{\text{Ga[bulk]}} + \mu_{\text{N[N}_2\text{]}} + \Delta H_f[\text{GaN}] \tag{67}$$

Since $\mu_{\text{N}} = \mu_{\text{N[N}_2\text{]}}$

$$\mu_{\text{Ga}} = \mu_{\text{Ga[bulk]}} + \Delta H_f[\text{GaN}] \tag{68}$$

i.e. in the N-rich limit the chemical potential of Ga will decrease[4] by the magnitude of the enhalpy of formation of GaN. In the Ga rich limit the chemical potential of Ga is given by Ga metal cohesive energy, whereas μ_{N} can be calculated via $\mu_{\mathrm{N}} = \mu_{\mathrm{N[N_2]}} + \Delta H_f[\mathrm{GaN}]$. For further discussion of chemical potentials, refer to [2].

The formation energy depends on the chemical potential of the electrons, the Fermi energy as it is often referred to[5]. Plotting this dependence for a variety of charge states allows one to determine thermodynamic transition levels, i.e. the Fermi levels at which the most stable charge state changes from one value to another [19], [20], [21].

In the above formation energy formalism, we are neglecting vibrational entropy contributions to the energy. It has been suggested that these effects are negligible for point defects [2]. At the moment a thorough computational evaluation including phonon frequencies, is computatationally too demanding.

3.8.2 Defect energy levels and electronic conductivity

In a perfect semiconductor system at zero temperature the electronic levels are formally occupied up to the Fermi energy, E_F, although in practice E_F lies in the bandgap where there are no electronic states. There are no electrons in the conduction band. By the band symmetry approximation, this *intrinsic* Fermi level lies near the middle of the energy gap. As the temperature increases, the Fermi function, giving the probability of an electron occupying an energy state E higher than the ground state, changes from a step function to a smeared out function as shown in Fig. 15. The number of electrons in the conduction band is then given by the integral of the probability over the electronic density of states in the conduction band. The

[4]Since the enthalpy of formation is exothermic the change to the chemical potential will be negative.

[5]The Fermi energy terminology here is potentially misleading. In a metal, the Fermi energy is the highest energy level that electrons will occupy at zero temperature. In semiconductors, this might not be the case as the Fermi energy may formally lie in the band gap and the highest energy electrons are below this level, in the valence band.

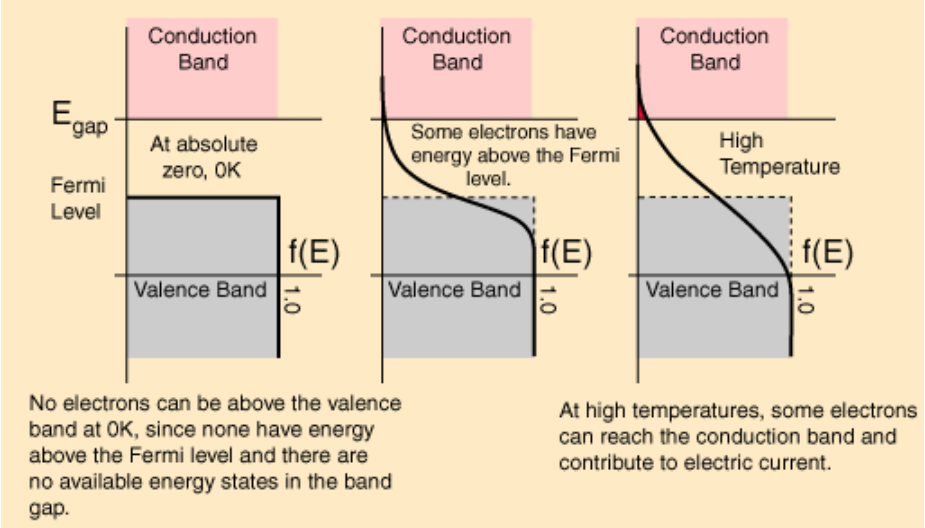


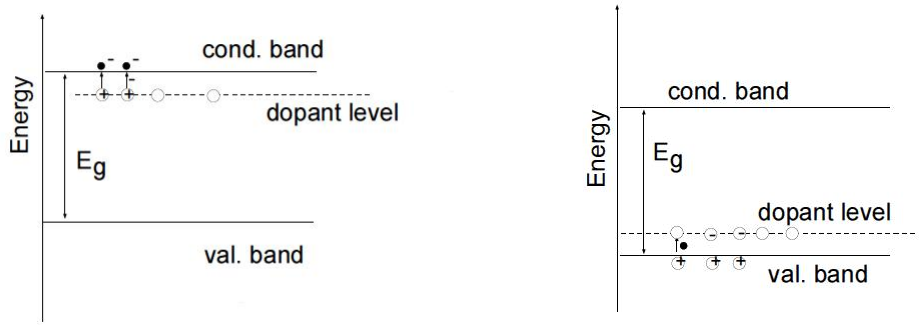
Figure 15: The Fermi function and temperature. From [23]

number of holes is determined by the corresponding integral in the valence band. The position of the Fermi level at non-zero temperature must be such that the number of electrons and holes is equal.

In semiconductors, the major source of electronic conductivity are defects. The defect energy levels often lie in the band gap as illustrated in Fig. 16. If the energy differences between the dopant level and the conduction or valence bands are small an electron can easily be removed from the valence band or added to the conduction band. The Fermi level shifts and hence so do the probabilities of the electron being found in the conduction band or a hole in the valence band. The number of free carriers increases and electrical conductivity rises with it.

A distinction is often made between shallow and deep defect levels. The energy difference between shallow defect levels and the VBM or CBM is small relative to the band gap. The bigger the difference the deeper the level. In general, shallow defect levels are associated with delocalised wavefunctions and small geometric distortions and vice versa for deep levels.

Initially the description "deep level" was applied to defects whose energies are found near the middle of the band gap. Many defects have properties that *cannot* be described by the same formalism as shallow defects and yet



(a) Donor doping (b) Acceptor doping

Figure 16: Donor and acceptor doping. The Fermi levels are shown by the dashed lines. E_g denotes the bandgap. From [24]

their energies are not in the vicinity of the middle of the band gap. These defects are now also called deep defects.

Shallow (hydrogenic) defects contain charge carriers (electrons or holes) that are only loosely bound to the defect centre, which is approximated as having infinite mass. A donor atom can be thought of as a hydrogen atom, but with electron screening which is determined by the electronic properties of the host system. Similarly an acceptor atom in this description is equivalent to a positron bound to a negative muon, with environment-dependent screening. Some of the important properties of these systems can be understood in terms of the effective mass approximation. These defects are most effectively studied by periodic DFT calculations, which are well-suited to their delocalised nature. Further details can be found in [25] and [7] .

The deep level defects are important in luminescence and lasing phenomena. They can also trap charge carriers and thus reduce electronic conductivity.

If a semiconductor system contains both shallow donors and acceptors, the holes created in the valence band and the electrons in the conduction band may recombine, decreasing free carrier concentrations, which is known as compensation.

Compensation and the deep/shallow nature of defects in a material determine the limits of conductivity of doped semiconductors. Other factors include the

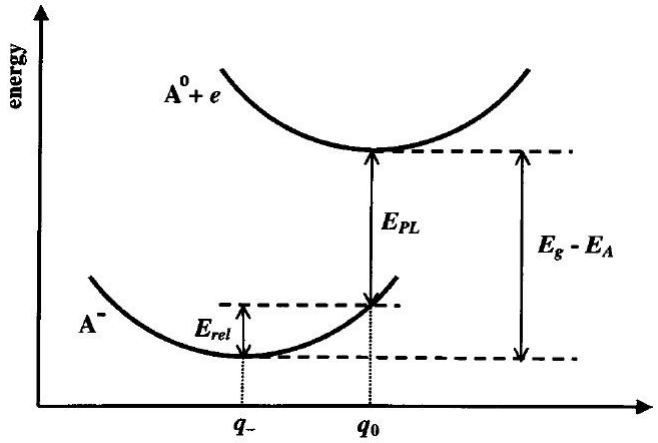


Figure 17: A configuration coordinate diagram showing the difference between thermodynamic and optical ionisation energies for an acceptor A. A^- denotes a defect that has trapped an electron whereas the state $A^0 + e^-$ corresponds to a neutral defect and an electron in the conduction band. $E_{\rm rel}$ is the Franck-Condon shift and is equal to the relaxation energy of the defect in the charge state A^- from its equilibrium geometry in the neutral state to its equilibrium geometry in the present charge state. E_{PL} is the energy measured in a photoluminescence measurement. The energy $E_g - E_A$ corresponds to the thermodynamic transition. From [34]

solubility of dopants.

Two types of defect level can be measured and calculated. The thermodynamic transition level between two charge states q_1 and q_2 is defined as the energy difference between the relaxed configurations of the two charge states, or, equivalently, the Fermi level where the two charge states have equal energies. This type of transition level is measured by deep-level transient spectroscopy (DLTS). For shallow centres it can be deduced from temperarature-dependent Hall measurements as the thermal ionisation energy.

The optical levels correspond to a transition where the geometry of the final charge state does not relax to its equilibrium configuration. These energies are determined in photoluminescence measurements. For a comparison of thermodynamic and optical levels see Fig. 17.

A relevant quantity, accessible to experiment, is the ionisation potential of the material, which is the difference between the Fermi level in the material and the vacuum level. The vacuum level, i.e. the energy of a stationary

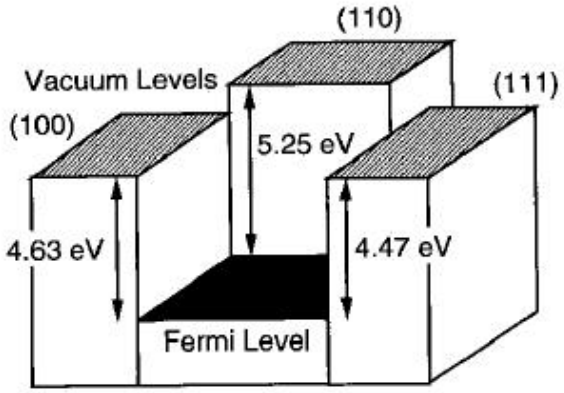


Figure 18: Vacuum levels for different tungsten surfaces from [26], values from [27].

electron infinitely removed from the solid, is, however, not uniquely defined as can be seen in Fig. 18, which shows the vacuum levels of tungsten as being different depending on which surface, (110), (111) or (100), the electron was removed from. The distinction is important in interpreting experimental measurements of the ionisation potential, especially in a highly ionic material such as GaN.

3.8.3 Defect geometries and electronic structure

There is no simple way of determining the geometry or electronic structure of defects without performing detailed calculations such as those presented later in this work.

Depending on the environment of the defect centre, the electronic structure can in the first instance be approximated with atomic orbitals, e.g. when considering a Ga interstitial or, alternatively, a molecular picture can be considered, e.g. for a nitrogen interstitial, which forms a bond with a lattice nitrogen in a fashion similar to a N_2 molecule.

In the atomic picture, the electronic structure of native defects can be understood with reference to s and p orbitals of the isolated Ga or N atoms. We will use the example of GaAs, which is discussed in more detail in [28] and shows many similarities to GaN, although it is most commonly found in the

zincblende phase and is less ionic than GaN. The energy levels of Ga vacancies in GaAs are shown in Fig. 19[6]. When an atom is removed, four dangling orbitals (ϕ_1, ϕ_2, ϕ_3 and ϕ_4) are formed, which point from the nearest neighbours to the now vacant site. A fully symmetric combination of these orbitals $\phi_1 + \phi_2 + \phi_3 + \phi_4$ is has a_1 symmetry (in group theory notation). The other three dangling orbitals belong to the t_2 representation and can be written as $(\phi_1 - \phi_2 - \phi_3 + \phi_4)$, $(\phi_1 + \phi_2 - \phi_3 - \phi_4)$ and $(-\phi_1 + \phi_2 - \phi_3 + \phi_4)$. These three orbitals are degenerate in the zincblende structure but split into a singlet and a doublet in the wurtzite structure due to the lowering of symmetry in wurtzite relative to zincblende. This is a consequence of the inequivalence of one of the nearest neighbours in wurtzite, which lies along the c axis from the vacancy, to the other three neighbours. In zincblende all four neighbours are equivalent. This approach has been taken for GaN in [38] and the hexagonal splitting can be seen in Fig. 26.

In open shell systems, electronic level degeneracy leads to symmetry-breaking and the lowering of energy according to the Jahn-Teller theorem. The associated geometric distortion removes the degeneracy, which is known as the Jahn-Teller effect. Pseudo-Jahn-Teller (or second order Jahn-Teller) effects have their origin in symmetry-breaking as well but it is the ionic energy that is reduced rather than electronic. This effect can occur in closed shell systems.

Certain defects in GaN are sometimes said to have negative-U properties. Such defects, usually associated with large geometry relaxations, undergo a thermodynamic transition with an associated change in the charge state of 2. For example, the thermodynamic transition may occur between the 1+ and 1- charge states while the charge 0 defect is not thermodynamically stable at any Fermi level.

[6]This approach is derived from the work of Coulson and Kearsley [29], which studied the energy levels of a vacancy in the perfect diamond lattice.

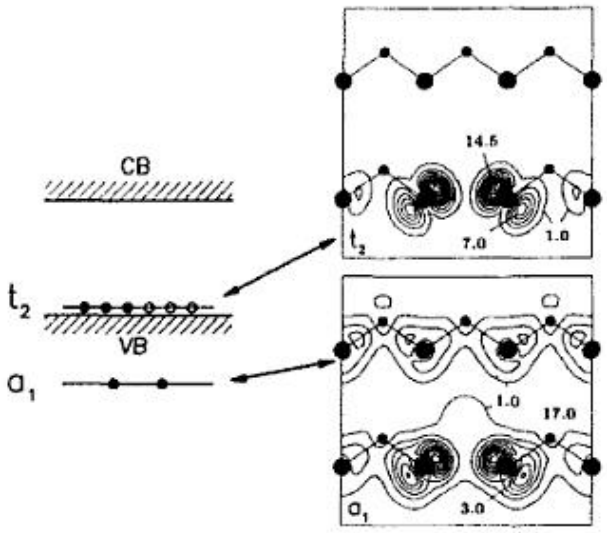


Figure 19: The energy level diagram and the electronic densities of the A1 singlet and the T2 triplet of a neutral Ga vacancy in ZB GaAs. It can be seen that in this material A1 is a resonance in the valence band whereas T2 is a bound state. Electron densities are in the (110) crystal plane. From [28].

3.9 Alloys

In this section, two important concepts relating to alloys are discussed, phase separation and cation ordering.

Phase separation is a phenomenon which is highly relevant for the applications of III-V nitrides. Alloys have many useful properties different from those of their constituent components, such as band gaps that result in the emission of different colours of light. If the solid solution is unstable with respect to decomposition to the binary compounds from which it was formed, these valuable properties cannot be accessed.

The second important concept discussed here is the ordering in an alloy and the enumeration of the ordered configurations in a computational study.

3.9.1 Phase separation

Alloys can under certain conditions decompose into their binary constituents (if the free energy of mixing is positive for all alloy compositions) or into

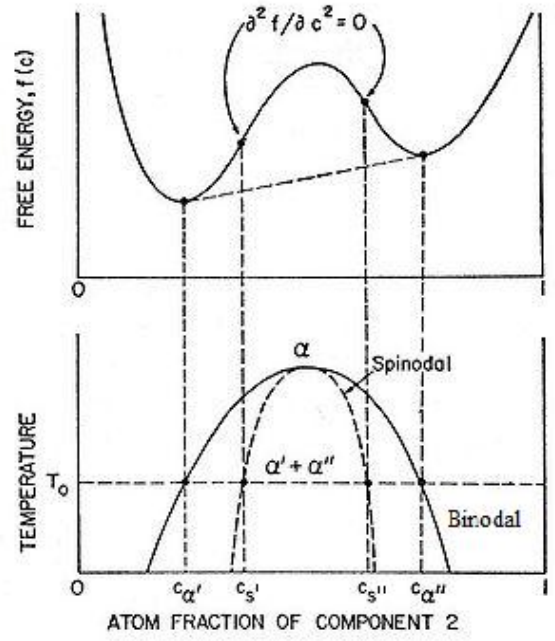


Figure 20: Free energy as a function of alloy composition (upper graph). The lower graph shows the spinode and binode as a function of temperature. Adapted from [31].

regions significantly richer or poorer in one constituent than would be expected from the overall composition of the alloy. The latter situation will arise if the free energy vs. composition has a dependence of the form shown in Fig. 20. For a system with such free energy characteristics, there are two types of decomposition, spinodal and binodal, determined by the curvature of the free energy vs. composition curve. Positive curvature leads to binodal decomposition and negative curvature to spinodal.

Spinodal decomposition is characterised by variations in composition that are small in amplitude but large in spatial extent. Such behaviour occurs because the spinodal region is unstable with respect to fluctuations in composition, which are small in their amplitude initially. The amplitude increases with time as can be seen in Fig. 21. Note that in spinodal decomposition the diffusion is uphill, i.e. against the concentration gradient.

Binodal decomposition occurs down the concentration gradient and is large

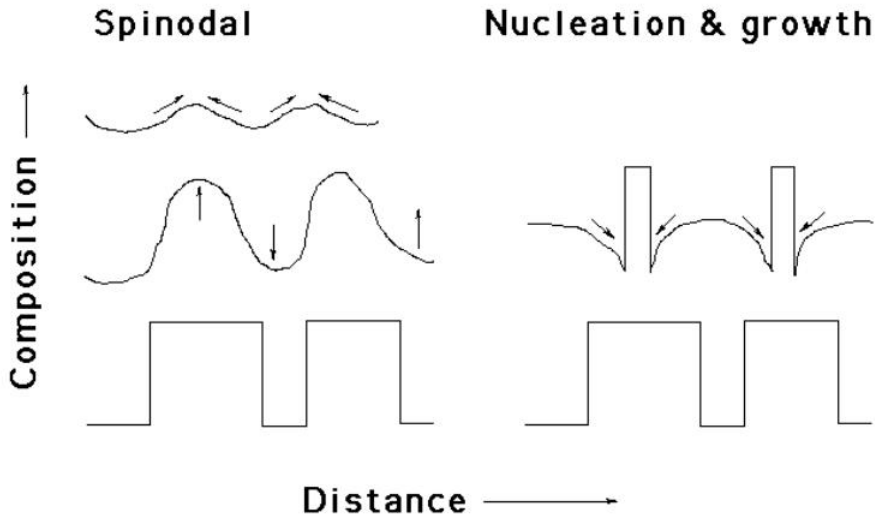


Figure 21: Spinodal vs. binodal decomposition. The arrows indicate the direction of particle flow. Time progression is from top to bottom. In the long time limit the concentration profiles converge. From [30].

in amplitude and small in extent. It occurs by nucleation and growth. The equilibrium concentrations are given by the positions of the minima of the free energy curve. The extent of A-rich and B-rich regions, where A and B are the constituent components, is determined by the conservation of A and B particle numbers. In contrast to spinodal decomposition, in nucleation and growth the compositions of the A and B rich regions do not change with with time but are fixed at the equilibrium concentrations from the start. The extent of the decomposition region grows with time, however. Additionally, there is a sharp concentration contrast between the regions with different compositions in nucleation and growth while in spinodal decomposition the concentration is more slowly varying in space.

In the limit of long times spinodal decomposition results in the same composition vs. spatial extent profile a binodal decomposition (see Fig 21).

For a system with a free energy vs. composition dependence like the one in the upper graph in Fig. 20, the two compositions where the curvature of the free energy vs. composition is zero, c_s' and c_s'', as well as the compositions corresponding to the minima of the free energy function, c_α' and c_α'', depend

on the temperature. The lower graph in Fig. 20 shows the dependence of these four points on temperature. The c_s points form the spinode and the c_α points are the binode curve on the lower graph in Fig. 20.

3.9.2 Ordering in alloys

If the alloy does not undergo phase separation, the lattice can be occupied randomly or there can be ordering in the arrangement. This section discusses the latter.

We model ordered structures using supercells (see Section 2.2.2 on p. 52), usually composed of a number of unit cells of the material. For example, a $2 \times 2 \times 2$ supercell is composed of two unit cells along each of the lattice vectors, giving eight unit cells in total. This supercell is then repeated an infinite number of times along each lattice vector to form the infinite crystal lattice. Once the structures have been generated, their energies can be calculated to determine the most thermodynamically stable arrangement.

In our study the anion sites are occupied by nitrogen ions. The cation sites may be occupied by Al, Ga or In. Different cation fractions in the unit cells lead to different macroscopic compositions of the alloy. Only ternary alloys of the form $A_xB_{1-x}N$, where A and B are cations, will be studied here.

We start with the smallest possible unit cell, equivalent to that of a wurtzite unit cell. The two cation sites may be occupied either by A or B. If both the sites are A or B, a binary compound is formed. If one of the cation sites contains A and the other one B, an ordered alloy $A_{0.5}B_{0.5}N$ results. Interchanging the two sites results in equivalent arrangements so in this case only one unique configuration is possible.

Next we consider a $2 \times 1 \times 1$ supercell. Now there are four cation sites. If we wish to create a 50:50 alloy, the number of possible arrangements is 4×3, some of which will be equivalent. The number of inequivalent sites will be determined by the symmetry of the system.

To establish whether two configurations are equivalent one can use the concept of isometric transformation. Such a transformation, e.g. rotation, translation, reflection, will do not change any of the distances and angles in the

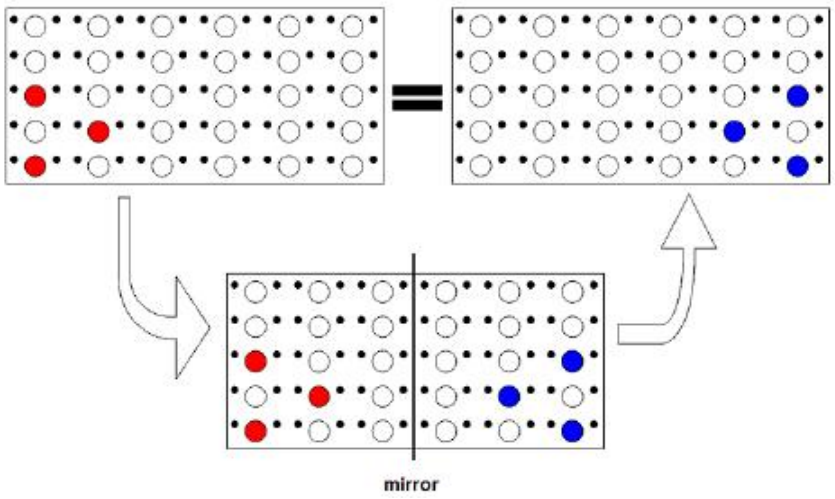


Figure 22: Symmetry operators of a parent structure (e.g. wurtzite) constitute the isometric operators of an alloy supercell. From [32]

final, transformed structure.

For a crystal structure, such as wurtzite, isometric operations are determined by the symmetry of the system. When these symmetry operations are applied to the binary compound (e.g. GaN), the structure is transformed back to itself. When the symmetry operators of the binary compound are applied to a supercell of an alloy material, it can be shown that equivalent structures are generated (see Fig. 22).

Clearly, for supercells containing more than a few atoms it would be extremely difficult to find all the inequivalent configurations so a software package, SOD (Site Occupancy Disorder) is used. This software utilises the symmetry operators of the relevant structure (wurtzite in our case) to establish whether a particular configuration is equivalent to any other configuration that has already been generated by the program. More details can be found in [32]. Since the original software did not contain the symmetry operators for the wurtzite structure, they were kindly implemented by Dr. Scott Woodley.

If a particular configuration, n, has a number of equivalent (degenerate) arrangements, M, each with energy E_n, then the probability of this configu-

ration is given by

$$P_n = \frac{1}{Z} M \exp\left(-E_n/k_B T\right) \tag{69}$$

Rearranging gives

$$P_n = \frac{1}{Z} \exp\left(-E_{\mathrm{red},n}/k_B T\right) \tag{70}$$

where k_B is the Boltzmann factor, T is the temperature and the reduced energy $E_{\mathrm{red},n}$ is defined as

$$E_{\mathrm{red},n} = E_n - T S_{\mathrm{deg},n} \tag{71}$$

The partition function, Z may be written as

$$Z = \sum_{n=1}^{N} \exp\left(-E_{\mathrm{red},n}/k_B T\right) \tag{72}$$

The degeneracy entropy, $S_{\mathrm{deg},n}$, is, by definition

$$S_{\mathrm{deg},n} = k_B \ln M \tag{73}$$

This contribution to the free energy is not to be confused with the vibrational entropy due to the phonon contributions, or with the entropy of the ensemble, i.e. its configurational entropy, S_{config}, discussed below.
The reduced energy and degeneracy entropy are not physically observable quantities; rather they are a mathematical convenience. The internal energy of the system, E, can be written as

$$E = \sum_{n=1}^{N} P_n E_n \tag{74}$$

The free energy can also be readily calculated as

$$F = -k_B T \ln Z \tag{75}$$

S_{config}, can then be deduced from

$$F = E - TS_{\text{config}} \tag{76}$$

Rearranging this equation and substituting for E from Eq.74 and F from Eq.75, we obtain

$$S_{\text{config}} = \frac{1}{T}\Big[\sum_{n=1}^{N} P_n E_n + k_B T \ln Z\Big] \tag{77}$$

Taking $M = 1$ for all configurations and using Eq. 69 to define E_n

$$S_{\text{config}} = \frac{1}{T}\Big[-\sum_{n=1}^{N} P_n k_B T \ln P_n Z + k_B T \ln Z\Big] = -k_B \sum_{n=1}^{N} P_n \ln P_n \tag{78}$$

Depending on the difference between the energies of the different configurations, the system can range from completely disordered (the energies are the same; all configurations equal probabilities) to ordered (one configuration is much more likely than any other). The maximum entropy, S_{max} in Eq. 78 will occur when all the configurations are equally likely, i.e. $P_n = 1/N$. In that case

$$S_{\text{max}} = k_B \ln N \tag{79}$$

N, the total number of possible configurations, can be determined for a given supercell size using combinatorial formulae. In the limit of infinite supercell size, it can be shown that for an alloy with composition $A_xB_{1-x}C$, the maximum configurational entropy, corresponding to a random alloy, is given by

$$S_{\text{config, max}} = -k_B[x \ln x + (1-x) \ln (1-x)] \tag{80}$$

References

[1] Zincblende-wurtzite polytypism in semiconductors, C.Y. Yeh, Z. W. Yu, S. Froyen, A. Zunger, Phys.Rev.B 46 **10086 - 10097** (1992)

[2] http://cst-www.nrl.navy.mil/lattice/struk/btype.html

[3] http://www.iue.tuwien.ac.at/phd/ayalew/node20.html

[4] http://kottan-labs.bgsu.edu/teaching/workshop2001/chapter5.htm

[5] http://www.fuw.edu.pl/~kkorona/interest.html

[6] Elastic constants and related properties of tetrahedrally bonded BN, AlN, GaN and InN, K.Kim, R.L.Lambrecht, B.Segall, Phys.Rev.B 53 **16310 - 16326** (1996)

[7] Fundamentals of semiconductors, P.Y.Yu, M.Cardona, Springer 2003

[8] The phonon dispersion of wurtzite-ZnO revisited, J.Serrano, F.J.Manjon, A.H.Romero, A.Ivanov, R.Lauck,M.Cardona, M.Krisch, Phys.Stat.Sol.(b) 244 **1478 - 1482** (2007)

[9] http://www.physik.uni-jena.de/~tief/Lehre/Optoelectronics

[10] Special points for Brillouin-zone integrations, H.J.Monkhorst, J.D.Pack, Phys.Rev.B 13 **5188 - 5192** (1976)

[11] Solid State Physics N.W.Ashcroft, N.D.Mermin, Holt-Saunders International Editions (1981)

[12] Band Structure and Reflectivity of GaN, S. Bloom, G. Harbeke, E. Meier, I. B. Ortenburger, physica status solidi (b) 66 **161 - 168** (1974)

[13] Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe, V.Bougrov , M.E.Levinshtein, S.L.Rumyantsev, A.Zubrilov, Eds. Levinshtein M.E., Rumyantsev S.L., Shur M.S., John Wiley & Sons, Inc., New York, (2001), 1-30.

[14] Optical properties of AlN epitaxial thin films in the vacuum ultraviolet region, H.Yamashita, K.Fukui, S.Misawa, S.Yoshida,J. Appl. Phys. 50 **896 - 898** (1979)

[15] Optical bandgap energy of wurtzite InN, T.Matsuoka, H.Okamoto, M.Nakao, H.Harima, E.Kurimoto, APL 81 **1246 - 1248** (2002)

[16] Band gap and band parameters of InN and GaN from quasiparticle energy calculations based on exact-exchange density-functional theory, P. Rinke, A. Qteish, M. Winkelnkemper, D. Bimberg, J. Neugebauer, M. Scheffler, APL 89 **161919** (2009)

[17] Chemical potential dependence of defect formation energies in GaAs - application to Ga self diffusion. S.B.Zhang, J.E.Northrup, Phys.Rev.B 67 **2339 - 2342** (1991)

[18] First-principles calculations for defects and impurities: Applications to III-nitrides, C. G. Van de Walle J. Neugebauer, J.Appl.Phys. 95 **3851 - 3879** (2004)

[19] Electronic structure, total energies and abundances of the elementary point defects in GaAs, G.A.Baraff, M.Schlüter, Phys.Rev.Lett. 55 **1327 - 1330** (1985)

[20] Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion, S.B.Zhang, J.E.Northrup, Phys.Rev.Lett. 67 **2339 - 2342** (1991)

[21] Formation energies and abundances of intrinsic point defects at the GaAs/AlAs(100) interface, M.Heinemann, M.Scheffler, Appl.Surface Sci. 56-58 **628 - 631** (1991)

[22] First-principles calculations for defects and impurities: Applications to III-nitrides, C.G.Van de Walle, J.Neugebauer, J.Appl.Phys.95 **3851 - 3879** (2004)

[23] `http://hyperphysics.phy-astr.gsu.edu/hbase/solids/fermi.html`

[24] `http://web.mit.edu/3.091/www/archives/Notes_3.pdf`

[25] Theory of defects in solids, A.M.Stoneham, Oxford University Press, 1975

[26] Energy Level Alignment and Interfacial Electronic Structures at Organic/Metal and Organic/Organic Interfaces, H.Ishii, K.Sugiyama, E.Ito, K.Seki, Adv. Mater. 11 **972 - 972** (1999)

[27] Work function measurements by the field emission retarding potential method, R. A. Strayer, W. Mackie, L. W. Swanson, Surf. Sci. 34 **225 - 248** (1973)

[28] Imperfections in III-V materials, E.R.Weber

[29] Colour centres in irradiated diamonds I, C.A.Coulson, M.J.Kearsley, Proc.R.Soc.London A 241 **433 - 454** (1957)

[30] `http://www.msm.cam.ac.uk/phase-trans/mphil/MP6-4.pdf`

[31] `http://en.wikipedia.org/wiki/Spinodal_decomposition`

[32] Symmetry-adapted configurational modelling of fractional site occupancy in solids, R.Grau-Crespo, S.Hamad, C.R.A.Catlow, N.H.de Leeuw, J. Phys.: Condens. Matter 19 **256201** (2007)

4 GaN, AlN and InN properties

This chapter focusses the properties of GaN; however, previous AlN and InN studies will be mentioned where appropriate. Many physical properties of III-nitrides have been studied exhaustively and a detailed discussion is beyond the scope of this work. Experimentally and computationally determined values of a number of these properties are referenced in Chapter 5, Tables 2 to 4 (bulk properties), 6 to 8 (Γ point phonon frequencies), 5 (wurtzite-to-rocksalt transition pressures) and Fig. 43 and 44 (thermal expansion data).
In the sections which will follow, we will review in more detail the literature on GaN interatomic potentials, previous QM/MM studies of solid state systems, calculations and relevant experimental data regarding point defects in GaN and the evidence for phase separation and ordering in alloys of III-V nitrides.

4.1 Interatomic potentials

A number of interatomic potentials have been developed for GaN employing a variety of functional forms. To the best of our knowledge, there are no previous formal charge models of GaN although such a model was developed for AlN [1].
[2] discusses a GaN model based on the Buckingham potential with Ga/N charges of 2+/2- and a shell model. Defect energies, phase transition pressures and surface properties were determined. [3] uses the potential derived in [2] to study the pressure dependence of phonon modes and thermodynamic stability of the different phases, including the less frequently discussed NiAs polymorph. [4] employs the same potential to study high pressure and temperature behaviour of GaN.
Another set of Buckingham potentials for the Al-Ga-In-N system, allowing the study of alloys was presented in [5]. Schottky and Frenkel defect energies as well as solution energies were calculated. The latter calculations indicate that In should readily dissolve in AlN and GaN at least in low concentrations, in contradiction to the widely discussed theoretical results in [6], which suggest they should be immiscible.

Yet another partial charge Buckingham model was developed in [7], this time using a genetic algorithm to find the potential parameters, rather than a Newton-Raphson process. It is somewhat less good at predicting the a structural parameter of wurtzite, calculating a value of 3.25Å compared to the experimentally determined 3.19Å. Ref. [8] presents a new partial charge Buckingham potential for Monte Carlo simulations of the thermodynamic properties of GaN/InN mixtures. Ref. [9] employs a potential which includes electrostatic energies due to bond and ionic charges to study wurtzite-zincblende stability. [10] presents a systematic approach for developing interatomic potentials for III-V semiconductors, validated for the case of AlAs, GaAs, and InAs. [11] then derives the AlN, GaN and InN potentials based on this methodology, which does not explicitly include a long range electrostatic interaction.

[12] presents a two body potential developed by fitting to Hartree-Fock data, which is used to model dislocations. Here charges of 1.15 and -1.15 are assigned to Ga and N, respectively. The short term interaction is described by Gilbert-type repulsion[7], a modified Morse potential and a van der Waals r^{-6} term.

In [13], a lattice inversion potential[8] for AlN, GaN and InN was developed. The ionic charges assigned to the Al/N are +0.98/-0.98, Ga/N +0.63/-0.63 and InN +0.82/-0.82. The short-range interaction is modelled with a Morse potential and an exponential repulsive term.

Other potentials, not based on the Buckingham function, include Stillinger-Weber [14], [15] [16] and bond-order (Tersoff-Brenner) [17] potentials, which do not explicitly include long range electrostatic forces. [17] and [14] were used in a study of native defects and the results compared with *ab initio* calculations. [16] were used to study detailed thermodynamic properties of GaN, including thermal expansion and specific heats. Finally, [18] discusses

[7]The Gilbert repulsion between ions i and j is of the form $(b_i + b_j)\exp\left(\frac{a_i+a_j-r_{ij}}{b_i+b_j}\right)$ where a_i is the repulsion radius of ion i and b_i is its softness parameter.

[8]A lattice inversion procedure employs DFT total-energy calculations for a number of polytypes, four in the case of [13], to derive the ionic charges and short range parameters of a MM model

a modified-embedded-atom model to study InGaN alloys.

4.2 QM/MM methodology

While QM/MM methods have been used extensively in enzyme, e.g. [19] and [20], and zeolites, e.g. in [21], [22] and [23], applications in ionic solids have been more sparse, and confined to oxides. These include studies of polar ZnO surface and their defects [1], methanol synthesis over Cu/ZnO [4], the interaction of ZnO surface sites with Cu atoms and clusters in various charge states [26]. MgO surfaces and their associated defects we discussed in [5]. The oxide studies listed here were carried out with the ChemShell software. Other codes include an early embedding software called ICECAP [2], using Hartree-Fock as the QM method, and PARAGAUSS [29], which was used to investigate Pd atom adsorption on the oxygen vacancies on the MgO surface [3]. All of these methodologies are based on the Mott-Littleton approach to defect calculations (see Section 2.1.2).
[31] reviews QM/MM methodology, the various implementations and their relative merits in more detail. The theoretical details are discussed in Section 2.3.

4.3 Defects

There are several outstanding questions regarding defects in GaN that computational studies have addressed. The formation energies of different types of defect have been studied by several workers and form the basis of much of the discussion on the following issues: the origin of n-type "autodoping" in the GaN and the difficulty in p-doping the material as well as limits to doping and compensation. The microscopic basis of yellow luminescence has also been the subject of much discussion. These issues are briefly summarised below.

4.3.1 Formation energies and relative abundance of defects

Interstitials and antisites have been reported in the literature as having higher formation energies than vacancies [32], which also reports nitrogen vacancies to be the most thermodynamically stable at the valence band maximum (VBM) and gallium vacancies at the conduction band minimum (CBM). Since GaN is most commonly n-type, the CBM value is of most interest. The calculations in [33] challenge the CBM results and find that in the Ga-rich limit at least the formation energies of Ga and N vacancies are comparable. While both calculations use periodic boundary conditions within the the LDA approximation (LSDA for [33]) they differ in their treatment of the Ga 3d electrons[9]. [33] uses the projector augmented wave method [36] whereas [32] employs the non-linear core correction (nlcc). The treatment of the Ga 3d electrons has a significant influence on the results obtained in periodic calculations. [33] studies the convergence of the defect energies reaching supercells of up to 300 atoms while [32] uses 96-atom supercells.
[37] reports the formation energies of defects in zincblende GaN and AlN calculated using the full potential linear muffin-tin orbital method. Vacancies are found to have lower formation energies in this polymorph as well.
There is significant variation in the calculated formation energies obtained in different studies, as can be seen from Fig. 23 and 24, and Table 35 in Section 8.3.5. The latter tabulates the formation energies of the neutral N vacancy calculated with various methods. Differences of up to a few eV are commonly observed.
The high concentration of Ga vacancies in n-type GaN is supported by positron annihilation experiments (PAS) [38] and the high concentrations were also correlated with increased intensity of yellow luminescence (see Section 4.3.2). The concentration was found to be particularly high in nitrogen-rich growth conditions, consistent with the theory in Section 3.8.1. Increasing

[9]There have been discussions over the role of the 3d electrons in the description of GaN. Periodic DFT calculations within the local density approximation have found that the 3d electrons play a role in bonding. Treating them as valence electrons is thought to improve the structural parameters and formation energies. However, their inclusion leads to a shrinking of the band gap in these calculations [34], [35].

the N:Ga ratio during the growth phase are associated with a decrease in free electrons, consistent with Ga acting as a compensating centre, according to the authors of [39]. In p-type GaN, Ga vacancies are lacking.
[34] argues that the lack of signal for negatively charged nitrogen vacancies in PAS measurements may be due to their signal being indistinguishable from that of the bulk. They also argue that the migration energies in the negative charge states of the N vacancy are low enough for equilibrium to be obtained on the timescale of the experiment, and lower than that of the Ga vacancy. Finally they suggest the low formation energies of N vacancies as the cause of the thermal decomposition of GaN at temperatures above about 900°C, which occurs via the evaporation of nitrogen from the surface.

The computational studies assume thermodynamic equilibrium, which may

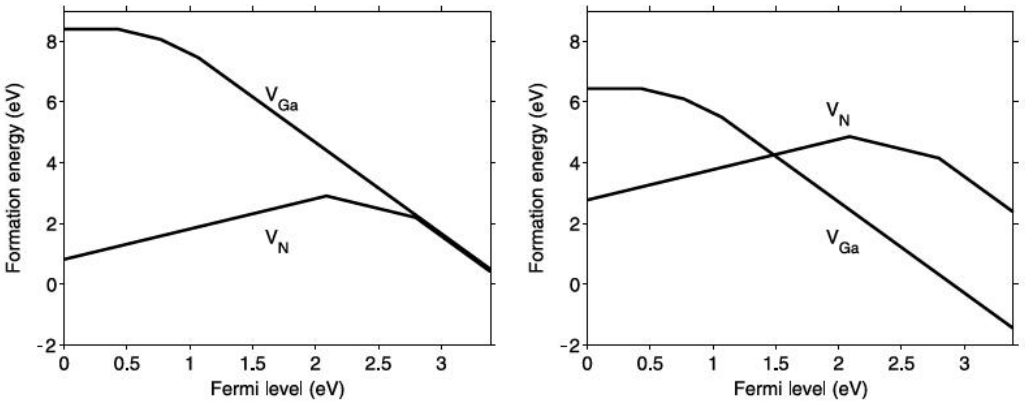


Figure 23: The formation energy of vacancies in GaN under Ga-rich (left) and N-rich (right) conditions, from [33]

or may not be achieved on the timescale of an experiment. The growth temperatures of the material may affect whether the concentrations of the defects equilibrate. MOCVD (Metal Organic Chemical Vapour Deposition) is carried out at relatively high temperatures (1000-1100° C). It is suggested that defect mobility is large enough for the equilibration of the concentrations [32], [34]. Molecular beam epitaxy (MBE) is typically performed at about 800° C and it is not clear whether thermodynamic equilibrium is attained.
Surface effects play an important role in the concentration of defects. Surfaces change the Fermi level in the band gap, and hence the formation energy

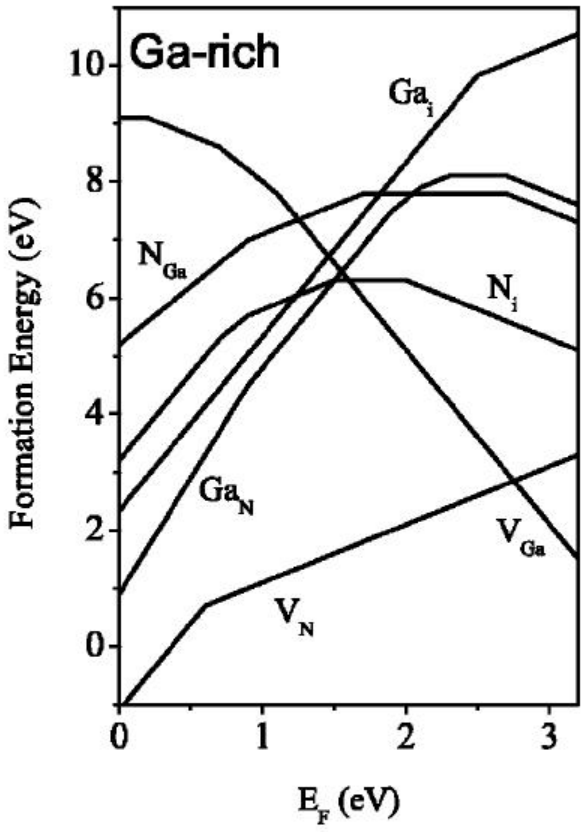


Figure 24: The formation energy of GaN defects under Ga-rich conditions using 96-atom supercells (for antisites 32-atom supercells were used) from [34].

of a defect. This often results in lower defect formation energies at the surface and consequently higher concentrations. Studies of defects at GaN surfaces can be found in e.g. in [42], [43], [44] and [45].

From the dependences of formation energies on the Fermi level the thermodynamic defect transition levels can be determined, as is shown in Fig. 25 from [34]. There is disagreement about the transition levels for the N vacancy. [34] suggests a 3+/1+ transition for at around 0.5eV (see Fig. 24) while [33] predicts 1+/1- and 1-/3- transitions further up in the band gap (see Fig. 23). The electronic structure of defects was calculated in [32]. Their results are shown in Fig. 26. The results are analysed in terms of zincblende symmetry with superimposed hexagonal splitting as has been previously applied to

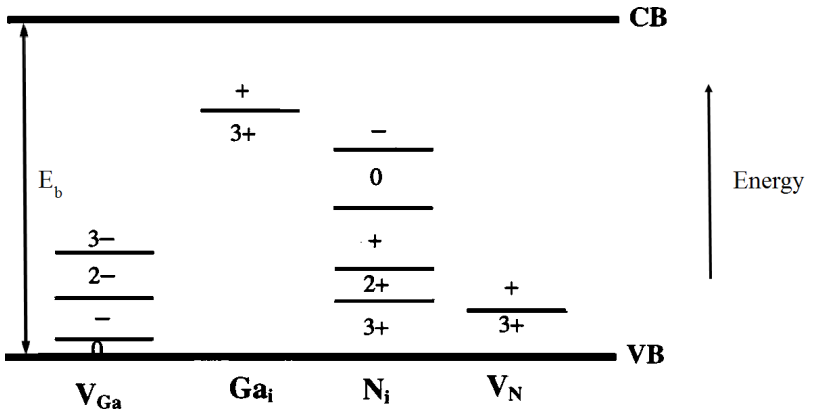


Figure 25: GaN thermodynamic defect transition levels, adapted from [34]. E_b is the band gap energy.

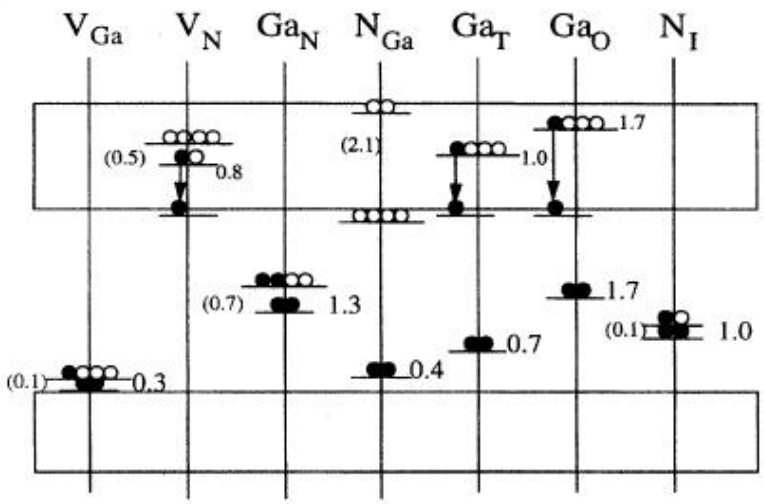


Figure 26: The electronic structure of neutral GaN native defects based on calculations in [38]. The T subscript refers to the tetrahedral arrangement and O refers to an octahedral arrangement.

GaAs studies (see Fig. 19). Hence the zincblende energy triplet level splits into a quasitriplet. This approach assumes predominantly covalent bonding. It is questionable how accurate such an analysis is for a highly ionic material such as GaN.

[32] studied the defect geometries and electronic structures using 32-atom supercells and periodic boundary conditions. They found that for a Ga va-

cancy the N dangling bonds are localised whilst for the N vacancy the Ga dangling bonds are strongly overlapping forming a metallic-like bond, which was also noted in [34]. They also find that with increasing occupation of the Ga vacancy defect levels from 3+ to 3- the first-neighbour N ions first move inward with the largest inward relaxation for the neutral charge state and then they move outwards in the case of negative charge states.

Tight-binding approaches use the fact that most Hamiltonian matrix elements are vanishingly small and hence only a sparse matrix needs to be diagonalised. Both empirical and *ab initio* descriptions have been attempted. However, due to the ionic nature of group-III nitrides the neglect of important interactions beyond the first neighbour leads to significant errors in defect energy levels and lattice constants. The energy levels of neutral nitrogen vacancies were predicted to be lie near to the bottom of the conduction band whereas periodic DFT predicts them to be near the top of the valence band [34].

[36] reports the formation energies and geometries of defects calculated with periodic DFT as well as Stillinger-Weber and Tersoff-Brenner potentials. The Ga interstitial is found to assume an unusual metastable split interstitial Ga-N configuration with DFT. The differences in formation energies between the different metastable Ga interstitials are found to be around 1eV at most.

Nitrogen antisite are studied in detail in [50]. The negative charge states experience a large Jahn-Teller effect.

Deep level transient spectroscopy (DLTS) and photoluminescence (PL) studies have been employed to study defect transition levels. PL measures optical transitions and is discussed in more detail in Section 4.3.2. DLTS measures the activation energy for the transition. A range of activation energies, only tentatively assigned to particular defects have been determined experimentally ranging from 0.18 to 0.62eV in references [31], [29] and [30]. Other studies by Haase et al. [41] have shown that a defect with activation energy 0.67eV in their experiment can be generated by N implanatation and then removed by annealing. They speculate that this points to the nitrogen vacancy or interstitial defect.

4.3.2 Luminescence in GaN

A whole range of luminescence lines have been observed in GaN, both undoped and intentionally doped. They are discussed in depth in [55]. Most of the lines in undoped GaN, e.g. blue, UV and red, are thought arise from defect complexes with impurities unintentionally introduced during the growth process. The yellow line is thought to be due to gallium vacancies or their complexes with impurities, and is discussed in more detail below.

The yellow luminescence (YL) band is centred around 2.2eV. It is observed in crystals as well as epitaxial layers prepared with a variety of techniques. It was suggested in [56] that YL is due to a shallow donor-deep acceptor transition. The proposed mechanism is shown in Fig. 27. The deep acceptor is 860meV above the valence band and in [56] the authors suggest it is a complex between a Ga vacancy and a C impurity. Other candidates that could give rise to yellow luminescence have since been advanced including surface defects and isolated Ga vacancy or the Ga vacancy-O_N complex. In [39] the shallow donor candidates are suggested to be Si_{Ga} and O_N. The deep acceptor is claimed to be a complex with a Ga vacancy.

An alternative mechanism has been proposed in [26] (see Fig. 28). The shallow donor D loses an electron to a deep double donor state DD in the charge state 1+, which is a non-radiative process (1). Radiative recombination (2) is postulated to occur next, with the deep donor state losing an electron to the acceptor state A, and a photon with energy 2.2eV is emitted. The direct D-A recombination competes weakly with the two-stage process. Evidence both in support and against this model is discussed in [55].

[55] presents the full experimental evidence related to YL in GaN. In this review, a gallium vacancy (or its complex) or a carbon impurity are suggested as the most likely causes of YL.

4.3.3 The role of defect in n-type vs. p-type doping of GaN

Historically there was some controversy about the origins of n-type doping in GaN. It was originally thought to be caused by the presence of a nitrogen

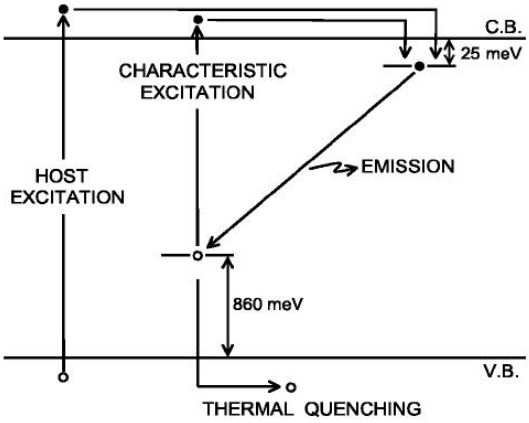


Figure 27: The suggested mechanism [56] for yellow luminescence in GaN.

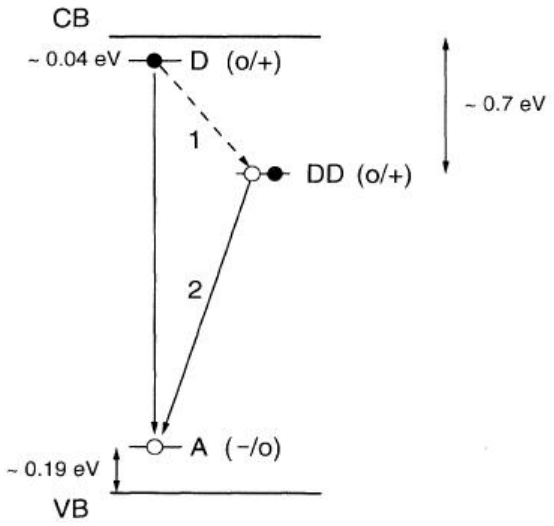


Figure 28: An alternative mechanism for the origin of yellow luminescence as proposed in [26].

vacancy but more recent research has attributed it to unintenional substitutional oxygen, replacing a N in the lattice. Oxygen as a source of n-type behaviour in GaN was first proposed in [58]. The authors cite the relationship between the carrier concentration and the growth temperature and the decrease in carrier concentration when the reactant NH_3 gas is purified to a high degree as evidence that impurities must be responsible for n-type be-

haviour of GaN.

A review of experimental evidence for the origins of n-type conductivity in GaN can be found in [59]. This includes evidence from high-pressure (>20GPa) freeze-out of electrons, which was originally attributed to the nitrogen vacancies creating a resonance in the conduction band under pressure. It is suggested in [59] that this behaviour is, however, also consistent with DX centre behaviour of an oxygen donor. A DX centre, such as a more widely studied Si in GaAs undergoes a strong relaxation under pressure accompanied by a transition from a shallow to deep centre. There is evidence for similar behaviour in experimental and in computational studies in AlGaN where an increase in Al content above about 30% will lead to oxygen becoming a deep donor [60] .

p-doped GaN was first made in 1989 by doping with Mg [61]. Increasing carrier concentrations in p-type doped GaN has proved challenging. Mg has been widely used as an acceptor but its solubility in GaN is limited. [59] suggests that the Mg defect with the lowest formation energy is a Mg substitution on a Ga site. H defects are thought to play an important role in p-doped GaN, where it acts as a donor, H^+, compensating an acceptor and passivating Mg, which leads to reduction in degree of p-doping. The formation of H complexes with Mg acceptors has been studied in [62] and on surfaces [63]. In the resulting complex H binds to N, rather than the neighbouring substitutional Mg and hence the vibrational frequency is more similar to that of a N-H bond.

Alternative dopants have been sought. Isoelectronic doping with In has been reported in [64]. Beryllium doping was studied in [65]. Fluorine has been proposed as a promising p-type dopant in AlN, ZnO and ZnMgO [66]. Calculations in [67] suggest that fluorine preferentially substitutes on N sites. Much research has been devoted to the causes of doping limitations in semiconductors, namely solubility of the dopant, its ionisation energy, incorporation of impurities in configurations which do not exhibit good doping behaviour and compensation either by native defects or impurities. For a fuller discussion see [34].

4.4 Alloys

Considerable attention has been devoted in the literature to the study of III-V alloys. Physical properties of the solid solutions have been widely studied, including the band gaps and piezoelectric properties, which are beyond the scope of this work. Here we consider the stability of the solutions with respect to phase separation and the extent of ordering in the alloys.

Early theoretical studies of $In_xGa_{1-x}N$ [6] suggested that the free energy of mixing should be positive and phase separation should occur, in apparent contradition with experiment which synthesised solid solutions of InGaN without any observations of phase separation, although it has proven difficult to synthesise InGaN with high In fractions. For example, [68] does not find any evidence of phase separation quantum wells with $In_xGa_{1-x}N$ layer thickness of 2.38 ± 0.10nm and $x = 0.18$ using a 3-dimensional atom probe (3DAP) and needle-shaped specimens. [69] reports $In_xGa_{1-x}N$ with x=0.35 without phase separation. [70] observes spinodal phase separation at $x = 0.29$ but not at lower In fractions. [71] reviews the experimental data regarding phase separation.

Other computational studies have reported similar results, with alloys with large lattice mismatch (AlInN and InGaN) being more prone to phase separation, with the binary compounds predicted to be immiscible up to temperatures of more than 1500°C [8]. AlN and GaN are well matched in their structural parameters and AlGaN is predicted to be stable at room temperatures [72], [18], [73], [74].

It has been suggested that the InGaN phase separation is suppressed in the interfacial region due to strain effects, which has indeed been observed in cross-sectional TEM images in [75]. Spinodal phase separation occurs in thick films far away from the interface, equivalent to the bulk environment. The critical thickness for the phase separation to be observed was found to be about 30-40nm [76].

The maximum film thickness of film that can be grown without phase separation was found to be dependent on the growth rate of the sample [77]. A high growth rate is required to produce homogenous InGaN alloys, implying

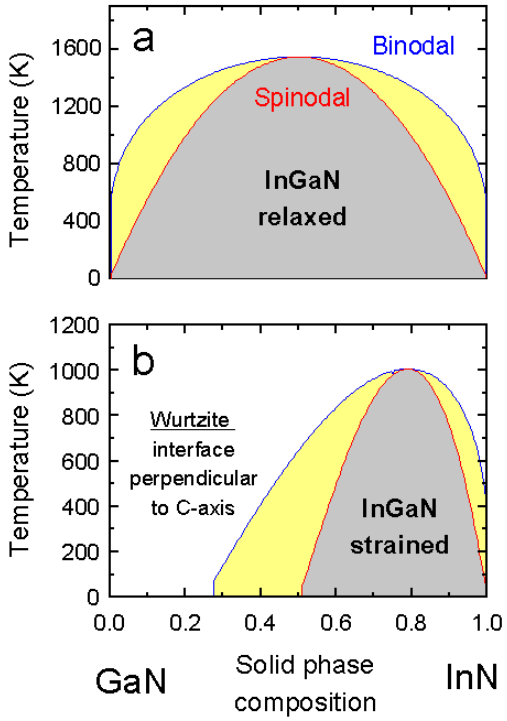


Figure 29: The phase diagram of $In_xGa_{1-x}N$ compounds for a) relaxed layers, b) strained layers with the interface in the ab plane. From[78] calculated with a valence force field model

that conditions far from thermodynamic equilibrium are required.

The suppression of phase separation by epitaxial strain has been confirmed theoretically in [78] using a valence field model. The spinodals and binodals are shown Fig. 29. The author also suggests that orienting the epitaxial layer parallel to the c axis will further stabilise the InGaN alloy (Fig. 30). In the case of the zincblende phase the tendency to phase separate is almost completely suppressed by the strain. [79] also finds evidence of suppression of phase separation in strained cubic InGaN alloys.

Another extensively discussed effect is "compositional pulling" [80]. Under the same growth conditions, alloys with different In fractions are observed depending on the substrate used in the region adjacent to the interface between the substrate and epitaxial layer. The molar fraction of In increases as the

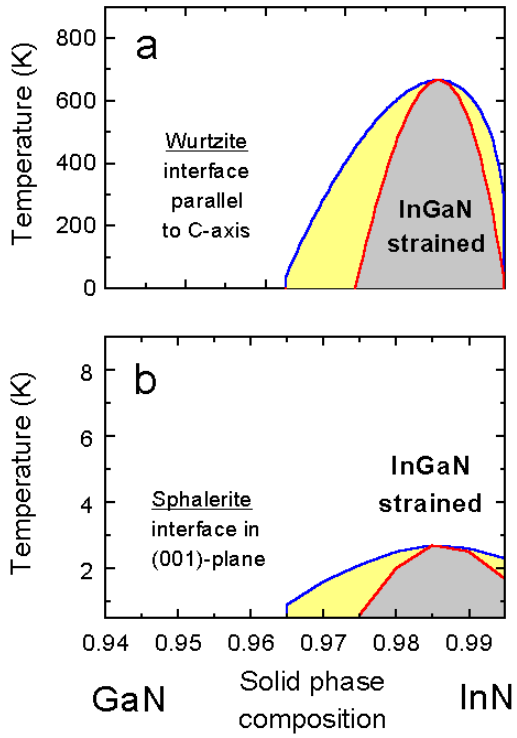


Figure 30: Phase diagram of $In_xGa_{1-x}N$ compounds for a) strained wurtzite layers parallel to the c-axis (b) strained zincblende (sphalerite) layers in (001)-plane. From [78]. The spinode and binode are shown.

thickness of the epitaxial layer increases. The greater the lattice mismatch between the substrate and the compound being deposited, the stronger the compositional pulling, which is attributed to the increasing strain. This phenomenon further complicates the relationship between strain and the degree of phase separation.

In summary, a variety of factors determine whether phase separation occurs in an alloy, including the degree of strain, growth conditions and the thickness of the epitaxial layer if the material is present as a film. It is possible that the resulting arrangements do not reflect the configurations at thermodynamic equilibrium.

[82] grew InGaN samples with In content $> 25\%$ and thickness $0.3-0.5\mu m$ at temperatures 690-780 ° C. Their results indicate a coexistence of phase sepa-

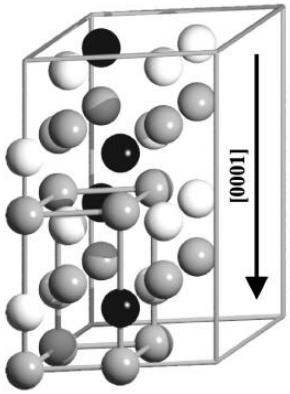


Figure 31: A lowest energy ordered $In_{0.25}Ga_{0.75}N$ structure from [81]. Ga atoms are white, In atoms are black and N are grey.

ration (areas with 96% InN by x-ray diffraction) and regions of ordering. Two types of ordering were observed, 1:1 ($In_{0.50}Ga_{0.50}N$) and 1:3 ($In_{0.25}Ga_{0.75}N$). The authors suggest ordering along the c-axis of -Ga-In-Ga-In- in the 1:1 case. The 3:1 case is postulated to have two types of high-symmetry ordering: either along the c-axis (Ga-Ga-Ga-In-Ga-Ga-Ga-In) or the following (in a $1 \times 1 \times 2$ supercell fractional coordinates): one Ga ion at (1/3, 2/3, 0), two Ga ions at (2/3, 1/3,1/4) and (2/3, 1/3, 3/4), and one In ion at (1/3, 2/3, 1/2).

[71] also discusses long range atomic ordering in InGaN based on experimental data, which relies on the appearance of "forbidden peaks" in the diffraction results. These peaks are thought to arise as a result of ordering of InN and GaN layers along the c-axis. Ordering is found to increase with film deposition rate. Further experimental evidence of such ordering is found by [69] (InGaN), [83] (AlGaN) and [84] (AlGaN). The ordering along the c-axis in wurtzite, with alternating layers of In/AlN and GaN is similar to the Cu-Pt type ordering found in zincblende III-V alloys [85]. This ordering was found in the computational study of InGaN using cluster expansion methods in [86].

Other geometries have also been reported. [87] finds evidence that the symmetry of an InGaN alloy is lowered to the P3ml spacegroup. [81] reports an ordered InGaN structure *without* layers of InN and GaN arranged along the

c-axis (Fig. 31) calculated using periodic DFT in the local density approximation.
In summary, there is some evidence of ordering in III-V nitride alloys, without a definitive answer as to the exact nature of the repeating unit.

References

[1] Intrinsic disorder in aluminum nitride A.N,Cormack, J.Am.Ceram.Soc.72 **1730 - 1732** (1989)

[2] An interatomic potential study of the properties of gallium nitride, P.Zapol, R.Pandey, J.D.Gale, J.Phys.: Condens.Matter 9, **9517 - 9525** (1997)

[3] Pressure effect on phonon modes in gallium nitride: a molecular dynamics study, M. R. Aouas, W. Sekkalb, A. Zaou, Solid State Communications 120 **413 - 418** (2001)

[4] Structural and thermodynamic properties of GaN at high pressures and high temperatures, X.Sun, Q.Chen, Y.Chua, C.Wang, Physica B: Condensed Matter 368 **243 - 250** (2005)

[5] Classical simulations of the properties of group-III nitrides, J A Chisholm, D W Lewis and P D Bristowe, J.Phys.Cond.Matt. 11 **L235 - L239** (1999)

[6] Solid phase immiscibility in GaInN, I.Ho and G. B. Stringfellow, Appl. Phys. Lett. 69 **2701 - 2703** (1996)

[7] Development of partial-charge potential for GaN, F.Gao, R.Devanathan, T.Oda, W.J. Weber, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 250 **50 -53** (2006)

[8] Monte Carlo simulation of GaN/InN mixtures, John A. Purton, Mikhail Yu. Lavrentiev, Neil L. Allan, J. Mater. Chem. 15 **785 - 790** (2005)

[9] An empirical potential approach to wurtzite-zincblende structural stability of semiconductors, T.Ito, Y.Kangawa, J.Cryst.Growth 235 **149 - 153** (2002)

[10] Systematic approach to developing empirical potentials for compound semiconductors, T.Ito, K. E. Khor, and S. Das Sarma , Phys.Rev.B 41 **3893 - 3896** (1990)

[11] Empirical interatomic potentials for nitride compound semiconductors, T.Ito, Jpn.J.Appl.Phys. 37 **L574 - L576** (1998)

[12] Molecular dynamics simulation of dislocations in wurtzite-type GaN crystal, K.Harafuji, T.Tsuchiya, K.Kawamura, J. Appl. Phys. 96 **2513 - 2524** (2004)

[13] Lattice inversion for interatomic potentials in AlN, GaN and InN, S.Zhang, N.Chen, Chem.Phys.309 **309 - 321** (2005)

[14] An empirical potential for the calculation of the atomic structure of extended defects in wurtzite GaN, N.Aichoune, V.Potin, P.Ruterana, A.Hairie, G.Nouet, E.Paumier, Comp.Mater.Sci.17 **380 - 383** (2000)

[15] A modifed empirical potential for energetic calculations of planar defects in GaN, J. Kioseoglou, H.M. Polatoglou, L. Lymperakis, G. Nouet,Ph. Komninou, Comp.Mat.Science 27 **43 - 49** (2003)

[16] Structural and thermodynamic properties of GaN: a molecular dynamics simulation, W.H.Moon, H.J.Hwang, Phys.Lett.A 315 **319 - 324** (2003)

[17] Modelling compound semiconductors: analytical bond-order potential for gallium, nitrogen and gallium nitride, J.Nord, K.Albe, P.Erhart, K.Nordlund, J.PHys.Condensed Matter 15 **035205** (2003)

[18] Atomistic modelling of III-V nitrides: modified embedded-atom method interatomic potentials for GaN, InN and $Ga_{1-x}In_xN$, E.C.Do, Y-H.Shin, B-J.Lee, J.Phys.Condens.Matt. 21 **325801** (2009)

[19] Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction, P.D. Lyne, A.J. Mulholland, W. G. Richards, J. Am. Chem. Soc. 117 (1995)

[20] QM/MM studies of enzymes, H.M.Senna, W.Thiel, Current Opinion in Chemical Biology 11 **11345 - 11350** (2007)

[21] Computer simulation of zeolite structure and reactivity using embedded cluster methods, P.Sherwood, A.H. de Vries, S.J. Collins, S.P.Greatbanks, N.A. Burton, M.A.Vincent, I.H. Hillier, Faraday Discuss. 106 **79 - 92** (1997)

[22] Stabilities of C3C5 alkoxide species inside H-FER zeolite: a hybrid QM/MM study, V,Nieminen, M.Sierka, D.Yu. Murzin, J.Sauer, Journal of Catalysis 231 **393 - 404** (2005)

[23] Lewis acidity and basicity of cation-exchanged zeolites: QM/MM and density functional studies, R.C.Deka, K.Hirao, J.Mol.Cat.A: Chemical 181 **275 - 282** (2002)

[24] Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials, A.A.Sokol, S.T.Bromley, S.A.French, C.R.A.Catlow, P.Sherwood, Int.J.of Quantum Chem. 99 **695 - 712** (2004)

[25] Identication and characterization of active sites and their catalytic processesthe Cu/ZnO methanol catalyst, S. A. French, A. A. Sokol, S. T. Bromley, C. R. A. Catlow, P. Sherwood, Topics in Catalysis 24 **161 - 172** (2003)

[26] Metal Cluster Support Interactions in the Cu/ZnO System: A QM/MM Study, S. T. Bromley, S. A. French, A. A. Sokol, C. R. A. Catlow, P. Sherwood,J. Phys. Chem. B 107 **7045 - 7057** (2003)

[27] Relative energies of surface and defect states: ab initio calculations for the MgO (001) surface, P.V.Sushko, A.L.Shluger, C.R.A.Catlow, Surface Sci. 450 **153 - 170** (2000)

[28] Hartree-Fock cluster computations of defect and perfect ionic crystal properties, J. H. Harding, A. H. Harker, P. B. Keegstra, R. Pandey, J. M. Vail and C. Woodward, Physica B+C 131 **151 - 156** (1985)

[29] Cluster embedding in an elastic polarizable environment: Density functional study of Pd atoms adsorbed at oxygen vacancies of MgO(001), V.A.Nasluzov, V.V.Rivanenkov, and A.B.Gordienko, K.M. Neyman, U.Birkenheuer, N.Rösch, J.Chem.Phys. 115 **8157 - 8171** (2001)

[30] QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis, P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjvoll, A. Fahmi, A. Schfer, Ch. Lennartz, J. Mol. Struct. (Theochem.) , 632 **1 - 28** (2003)

[31] QM/MM: what have we learned, where are we, and where do we go from here?, H.Lin, D.G.Truhlar, Theor.Chem.Acc 117 **185 - 199** (2007)

[32] Diffusivity of native defects in GaN, S.Limpijumnong, C.G.van de Walle, Phys.Rev.B 69 **035207** (2004)

[33] Vacancies in wurtzite GaN and AlN, K. Laaksonen, M. G. Ganchenkova, R. M. Nieminen, J. Phys. Condens. Matter 21 **015803** (2009)

[34] First-principles calculations for defects and impurities: Applications to III-nitrides, C. G. Van de Walle J. Neugebauer, J.Appl.Phys. 95 **3851 - 3879** (2004)

[35] Native defects and impurities in cubic and wurtzite GaN, J.van de Walle, MRS **687 - 692** (Spring 1994)

[36] Projector augmented-wave method, P. E. Blöchl, Phys. Rev. B 50 **17953 - 17979** (1994)

[37] Calculated defect levels in GaN and AlN and their pressure coefficients, I. Gorczyca, A. Svane, N. E. Christensen, Solid State Communications 101 **747 - 752** (1997)

[38] Observation of Native Ga Vacancies in GaN by Positron Annihilation, K. Saarinen, T. Laine, S. Kuisma, J. Nissil, P. Hautojrvi, L. Dobrzynski, J. M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszczynski, I. Grzegory, and S. Porowski, APL 79 **3030 - 3033** (1997)

[39] Gallium vacancies and the growth stoichiometry of GaN studied by positron annihilation spectroscopy, K.Saarinen, P.Seppala, J.Oila, P.Hautojarvi, C.Corbel, O.Briot, R.L.Aulombard, APL 73 **3253 - 3255** (1998)

[40] Nitrogen vacancies as major point defects in gallium nitride, M.G.Ganchenkova, R.M.Nieminen, PRL 96 **196402** (2006)

[41] Luminescence properties of defects in GaN, M.A.Reshchikov, H.Horkoc, J.Appl.Phys.97 **061301** (2005)

[42] Surface sensitivity of impurity incorporation: Mg at GaN (0001) surfaces, C.Bungaro, K. Rapcewicz, J. Bernholc, Phys.Rev.B **9771 - 9774** (1999)

[43] Adatom diffusion at GaN (0001) and (0001) surfaces, T.Zywietz, J.Neugebauer, M.Scheffler Appl.Phys.Lett.73 **487 - 489** (1998)

[44] Adsorption and incorporation of silicon at GaN(0001) surfaces, A. L. Rosa, J. Neugebauer, J. E. Northrup, Chae-Deok Lee, R. M. Feenstra, Appl. Phys. Lett. 80 **2008 - 2010** (2002)

[45] Review of Structure of Bare and Adsorbate-Covered GaN(0001) Surfaces, R.M. Feenstra, J. E. Northrup, J. Neugebauer, MRS Internet Journal Nitride Semiconductor Research **3** (2002)

[46] Point-defect complexes and broadband luminescence in GaN and AlN, T. Mattila, R. M. Nieminen, Phys. Rev. B 55 **9571 - 9576** (1997)

[47] Atomic geometry and electronic structure of native defects in GaN, J.Neugebauer, C.G. van de Walle, Phys.Rev.B 50 **8067 - 8070** (1994)

[48] Native defects in Gallium Nitride, R.Boguslawski, E.L.Briggs, J.Bernholc, Phys.Rev.B 51 **17255 - 17258** (1995)

[49] Intrinsic defect properties in GaN calculated by *ab initio* and empirical potential methods, F.Gao, E.J. Bylaska, W. J. Weber, Phys.Rev.B **245208** (2004)

[50] *Ab initio* study of Nitrogen antisites in GaN and AlN, T. Mattila, R. M. Nieminen, A.P.Seitsonen, Proceedings of the III-V nitride Symposium at the 189th ECS meeting, Los Angeles **205 - 211** (1996)

[51] Deep level defects in n-type GaN, W. Götz, N. M. Johnson, H. Amano, I. Akasaki, Appl. Phys. Lett. 65 **463 - 465** (1994)

[52] Deep centers in n-GaN grown by reactive molecular beam epitaxy, Z-Q. Fang, D. C. Look, W. Kim, Z. Fan, A. Botchkarev, H. Morkoc, APL 72 **2277 - 2279** (1998)

[53] Deep levels in the upper band-gap region of lightly Mg-doped GaN, P. Hacke, H. Nakayama, T. Detchprohm, K. Hiramatsu, N. Sawaki, APL 68 **1362** (1996)

[54] Deep-level defects and n-type-carrier concentration in nitrogen implanted GaN, D. Haase, M. Schmid, W. K ürner, A. Dörnen, V. Härle, F. Scholz, M. Burkard, H. Schweizer, APL 69 **2525** (1996)

[55] Luminescence properties of defects in GaN, M.A.Reshchikov, H.Morkoc, J.Appl.Phys. 97 **061301** (2005)

[56] Mechanism of Yellow Luminescence in GaN, T.Ogino, M.Aoki, Jpn. J. Appl. Phys. 19 **2395 - 2405** (1980)

[57] Optically detected magnetic resonance of GaN films grown by organometallic chemical vapour deposition, E.R.Glaser, T.A.Kennedy, K.Doverspike, L.B.Rowland, D.K.Gaskill, J.A.Freitas Jr., M.Asif Khan, D.T.Olson, J.N.Kuznia, D.K.Wickenden, Phys.Rev.B 51 **13326 - 13336** (1995)

[58] On the origin of free carriers in high-conducting n-GaN, W. Seifert, R. Franzheld,Dr. E. Butter, H. Sobotta, V. Riede, Crystal Research and Technology 18 **383 - 390** (1983)

[59] Theory of doping and defects in III-V nitrides, C.G.van de Walle, C.Stampfl, J.Neugebauer, J.Cryst.Growth 189-190 **505 - 510** (1998)

[60] Metastability of Oxygen Donors in AlGaN, M. D. McCluskey, N. M. Johnson, C. G. Van de Walle, D. P. Bour, M. Kneissl, W. Walukiewicz , Phys. Rev. Lett. 80 **4008 - 4011** (1998)

[61] P-Type Conduction in Mg-Doped GaN Treated with Low-Energy Electron Beam Irradiation (LEEBI), H.Amano, M.Kito, K.Hiramatsu, I.Akasaki, Jpn. J. Appl. Phys. 28 **L2112 - L2114** (1989)

[62] Hydrogen in GaN: Novel Aspects of a Common Impurity, J.Neugebauer, C. G. Van de Walle, Phys. Rev. Lett. 75 **4452 - 4455** (1995)

[63] First-Principles Surface Phase Diagram for Hydrogen on GaN Surfaces, C.G. Van de Walle, J. Neugebauer, Phys.Rev.Lett. 88 **066103** (2002)

[64] Characteristics of p-type GaN Films Doped with Isoelectronic Indium Atoms, F. C. Chang, K. C. Shen, H. M. Chung, M. C. Lee, W. H. Chen, W. K. Chen, CHINESE JOURNAL OF PHYSICS VOL. 40 **637 - 643** (2002)

[65] P-type doping of GaN, R. K. Wong,M.S. Thesis (2000)

[66] A pathway to p-type wide-band-gap semiconductors, A. Janotti, E.Snow, C. G. Van de Walle, Appl. Phys. Lett. 95 **172109** (2009)

[67] First-principles calculations for effects of Fluorine impurity in GaN, J. Lu, M. Gao, J. Zhang, Y. Wang, Z.Yu `http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04648280`

[68] Does In form In-rich clusters in InGaN wells?, C.J.Humphreys, Philosophical Magazine 87 **1971 - 1982** (2007)

[69] Phase separation and ordering in InGaN alloys grown by molecular beam epitaxy, D. Doppalapudi, S. N. Basu, K. F. Ludwig, Jr., T. D. Moustakas, J.Appl.Phys. 84 **1389 - 1395** (1998)

[70] Microstructural characterization of $In_xGa_{1-x}N$ MBE samples, M.Katsikini, E.C.Paloura, F.Boscherini, F.D'Acapito, C.B.Lioutas, D.Doppaludi, Nuclear instruments and methods in physics research B 200 **114 - 119** (2003)

[71] Article "Study of Phase Separation and Ordering in InGaN Alloys" by S.N.Basu in III-V nitride materials and processes III By T. D. Moustakas, S. E. Mohney, S. J. Pearton, Electrochemical Society. Dielectric Science and Technology Division, Electrochemical Society

[72] Molecular simulation study of miscibility of ternary and quaternary InGaAlN alloys, J.Adhikari, D.A.Kofke, J.Appl.Phys.95 **6129 - 6137** (2004)

[73] Single phase In_xGa_{1xN} (0.25 x 0.63) alloys synthesized by metal organic chemical vapor deposition B.N. Pantha, J. Li, J.Y. Lin, H. X. Jiang, Appl.Phys.Lett. 93 **182107** (2008)

[74] Theoretical study of phase separation in wurtzite InGaN, J.Zheng, J.Kang, Mat.Sci.in Semicond.Process. 9 **341 - 344** (2006)

[75] Compositional dependence of phase separation in InGaN layers, M.Rao, D.Kim,S.Mahajan, Appl.Phys.Lett.85 **44 - 51** (2004)

[76] A Novel Approach to the Direct, Nanoscale Compositional Analysis of InGaN Layers by means of Laser Pulsed-Atom Probe Tomography, M.Müller, B. Gault, D. Saxey, A. Cerezo, G. Smith, ISCS 2010 conference

[77] Evolution of phase separation in In-rich InGaN alloys, B. N. Pantha, J. Li, J. Y. Lin, H. X. Jiang, Appl. Phys. Lett. 96 **052110** (2010)

[78] Suppression of phase separation in lnGaN due to elastic strain, Karpov, S Yu, MRS Internet Journal of Nitride Semiconductor Research. Vol. 3 textbf16 (1998)

[79] First-principles calculations of the thermodynamic and structural properties of strained $In_xGa1-xN$ and $Al_xGa1-xN$ alloys, L.K.Teles, J.Furthmüller, L.M.R.Scolfaro, J.R.Leite, F.Bechstedt, Phys.Rev.B 62 **2475 - 2485** (2000)

[80] The Composition Pulling Effect in MOVPE Grown InGaN on GaN and AlGaN and its TEM Characterization, K. Hiramatsu, MRS , `http://www.mrs.org/s_mrs/doc.asp?CID=26415&DID=321056`

[81] Limits and accuracy of valence field models for $In_xGa_{1-x}N$ alloys, F.Grosse, J.Neugebauer, Phys.Rev.B 63 **085207** (2001)

[82] Phase separation and ordering coexisting in InxGa1-xN grown by metal organic chemical vapor deposition, M. K. Behbehani, E. L. Piner, S. X. Liu, N. A. El-Masry, S. M. Bedair, Appl.Phys.Lett.75 **2202 - 2204** (1999)

[83] Long range order in $Al_xGa_{1-x}N$ films grown by molecular beam epitaxy, D. Korakakis, K. F. Ludwig, Jr., and T. D. Moustakas, Appl. Phys. Lett. 71, 72 **72 - 74** (1997)

[84] Chemical ordering in AlGaN alloys grown by molecular beam epitaxy, E. Iliopoulos, K. F. Ludwig, T. D. Moustakas, S. N. G. Chu, Appl. Phys. Lett. 78 **463 - 465** (2001)

[85] Handbook on semiconductors Vol.3, A.Zunger, S.Mahajan, North-Holland, Amsterdam (1994)

[86] Phase separation and ordering in group-III nitride alloys, L.K.Teles, M.Marques, L.M.R.Scalforo, J.R.Leite, L.G.Ferreira, Braz.J.Phys.34 **593 - 597** (2004)

[87] Chemical ordering in wurtzite $In_xGa_{1-x}N$ layers grown on (0001) sapphire by metalorganic vapour phase epitaxy, P.Ruterana. G.Nouet, W.Van der Stricht, I.Moerman, L.Considine, APL 72 **1742 - 1744** (1998)

5 Molecular Mechanics studies of III-V nitride semiconductors

In this section new formal charge interatomic potentials for AlN, GaN and InN are presented. The extra constraint of formal charges makes fitting such potentials a challenging problem due to the degrees of freedom lost in restricting the ionic charges. To add extra degrees of freedom, independently parameterised layers of short range interactions are used in the model, in contrast to previous models for III-V nitrides, which are single-layered.
Since the potentials have a common model for the interactions between N ions (apart from InN having a different N spring constant) they can be extended to the study of alloys.
The potentials were rigorously tested by performing a number of calculations of physical properties. In addition to the bulk properties of the wurtzite, zincblende and rocksalt phases, some of which were used in the fitting process, the temperature and pressure behaviour are compared with experiment and other theoretical studies. These calculations go beyond the usual validation tests to which most potential models are often subjected.

5.1 Methodology

The General Utility Lattice Program (GULP) package [1], [2] was used for fitting interatomic potentials and calculating physical properties in this work. The theoretical background is discussed in Chapter 2. The software is suitable for fitting the potential parameters as well as for structural optimisation (energy minimisation). These functionalities are described in more detail in Chapter 2. The methods of calculation of bulk properties, such as elastic constants, from the energy expression are reviewed in Chapter 3.
The starting point for the development of the GaN potential was the partial charge GaN model of Zapol et al. [3]. The partial charges were increased in small increments until a formal charge model was obtained. The effects of the N core/shell charge split were then studied. Whilst GULP has a facility for

fitting parameters, the process is highly manual as the starting parameters in a fitting procedure must be sufficiently close to the optimum parameterisation for a successful fit.
Furthermore, the correct relative thermodynamic stability of different polymorphs is a difficult observable to include in fitting and it has to be tracked manually. The fitting of phonon frequencies is also problematic. While it is possible to fit to a numerical value of a frequency, there is no guarantee that this will be assigned to the correct mode as the order of modes might not be correctly reproduced by the model.
The parameterisation was further refined once it was used in a QM/MM scheme as the QM and MM forces were mismatched. This problem is related to an inherent difficulty with interatomic potentials as the errors in interatomic interactions can compensate for each other and lead to seemingly correct results. In this case, improving the model based on the QM information also led to better predictions of piezoelectric constants. The consistency of the final GaN potential with QM calculations gives credence to the model.
Based on the GaN parameters InN and AlN parameters were fitted. The InN dielectric constants were very challenging to reproduce while keeping the same spring constant as in GaN and therefore it was necessary to fit k_2 and k_4 constants of the shell model (see Section 2.1) separately for InN, which presents some difficulties when modelling alloys.
The model is based on the commonly used short range interatomic interactions, such as Buckingham and Born-Meyer. However, the parameters of these interactions are fitted separately for different atomic neighbour shells unlike in previous studies where the parameters are the same for the full range of interatomic distances. The experimental near-neighbour interatomic distances in the materials studied, which are used as a guide to delimiting the parameterisation layers in our model, are tabulated in Appendix I.
The successive layers of the usual interatomic potential functions are joined by polynomial functions designed to ensure smoothness (see Fig. 34 for illustration). The extra parameters in the present model compared to a single layer potential add degrees of freedom. They should in principle lead to a better fit and compensate for the extra constraint of fixed ionic charges.

The connecting polynomial functions are of order five in the interatomic distance r and are chosen to match the end-point functions in the value of the potential and its first and second derivatives

$$V_{\text{polynomial}} = c_5 r^5 + c_4 r^4 + c_3 r^3 + c_2 r^2 + c_1 r + c_0 \tag{81}$$

The following conditions are imposed on the polynomial acting between r_1 and r_2, the end points of the polynomial region:

$$V_1(r_1) = V_{\text{polynomial}}(r_1) \text{ and } V_2(r_2) = V_{\text{polynomial}}(r_2) \tag{82}$$

$$\left(\frac{dV_1}{dr}\right)_{r_1} = \left(\frac{dV_{\text{polynomial}}}{dr}\right)_{r_1} \text{ and } \left(\frac{dV_2}{dr}\right)_{r_2} = \left(\frac{dV_{\text{polynomial}}}{dr}\right)_{r_2} \tag{83}$$

$$\left(\frac{d^2V_1}{dr^2}\right)_{r_1} = \left(\frac{d^2V_{\text{polynomial}}}{dr^2}\right)_{r_1} \text{ and } \left(\frac{d^2V_2}{dr^2}\right)_{r_2} = \left(\frac{d^2V_{\text{polynomial}}}{dr^2}\right)_{r_2} \tag{84}$$

where potential V_1 acts at interatomic distances $< r_1$ and V_2 is valid $> r_2$. V_1 and V_2 are traditional potentials, such as the Buckingham or other potentials discussed in Section 2.1.

The number of layers is varied as necessary. In some cases the potentials have the same parameterisation for the full range of interatomic distances, as is commonly used in MM studies. In other cases up to three different layers with varying parameterisations, or, indeed, different functional forms, are present and connected with polynomial regions.

The polynomial functions act at interatomic distances at which ions are not observed in the bulk material and hence do not affect bulk properties. When the material is placed under pressure or a defect is introduced, ions might enter the polynomial regions. However, by ensuring the smoothness of the polynomials relative to the end-point functions, it is assumed that even the polynomial regions model the physical properties of the system reasonably well.

A final requirement for an acceptable potential function is that it must not

exhibit any unphysical maxima or minima.
The potentials are taken to act on the cores of Ga, In and Al and the shells of N ions. In the case of N ions, both the shells and the cores experience forces as shown in the parameter listings in Appendix I. As the shells are attached to the cores by springs, the forces acting on the shells are largely transmitted to the cores and vice versa. Relative shell-core displacement within the shell model (Section 2.1) simulates electronic polarisation effects in the material. We aimed to fit the lattice parameters to within 1% of experimental data; the dielectric constants were expected to be within 10%. Close to the energy minimum the high frequency dielectric constants are relatively simple to fit as they depend almost exclusively on the values of spring constant and shell charge. Since the elastic constants have large experimental errors less stringent requirements were imposed. Zone-centre phonon frequencies were also fitted.
All the energy calculations at this stage were performed without taking into account phonon vibrations. Internal lattice energies, rather than free energies, were thus minimised. Including vibrations in a quantitatively correct fashion is a demanding task for interatomic potentials unless they are specifically parameterised for that purpose.

5.1.1 Molecular mechanics model for GaN, InN and AlN

The final potential parameters for GaN, InN and AlN are presented in Tables 45, 47 and 48 in Appendix I, respectively.
The N-N short range potential is common to all three materials. It contains a region of Morse potential (see Section 2.1), which is more commonly employed to model covalent species rather than a strongly ionic material. However, it has been found empirically to work well in our models. The short range potential has a tail composed of a r^{-6} function. The short range potential only is plotted in Fig. 32, where the different layers are distinguished by colour, and the short range and long range contributions are contrasted in Fig. 33. Similarly the Ga-N potential is depicted in Fig. 34 and 35.

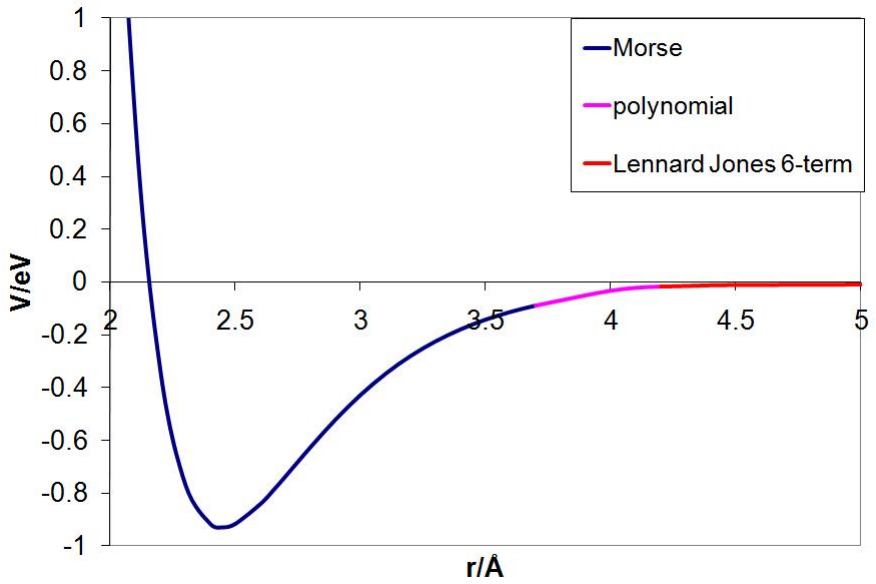


Figure 32: The N-N short range potential.

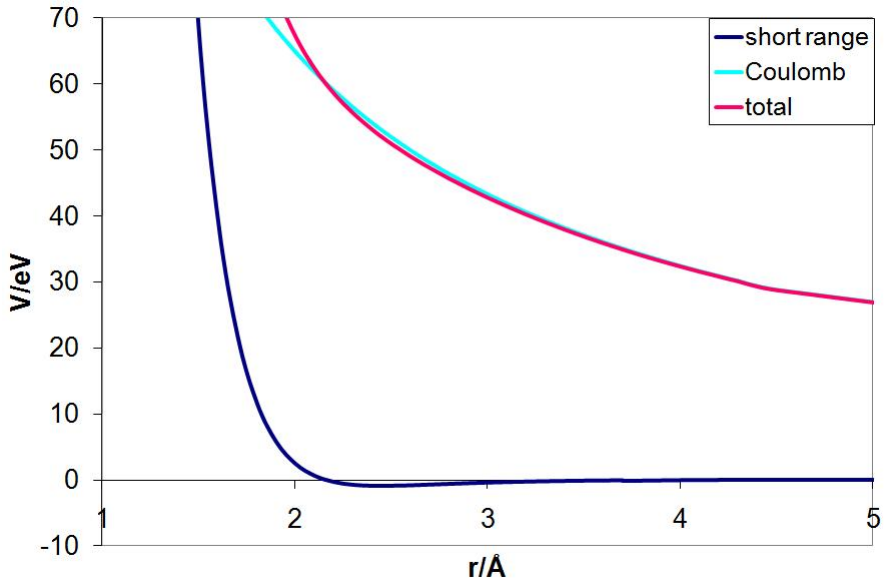


Figure 33: The N-N total potential and its components.

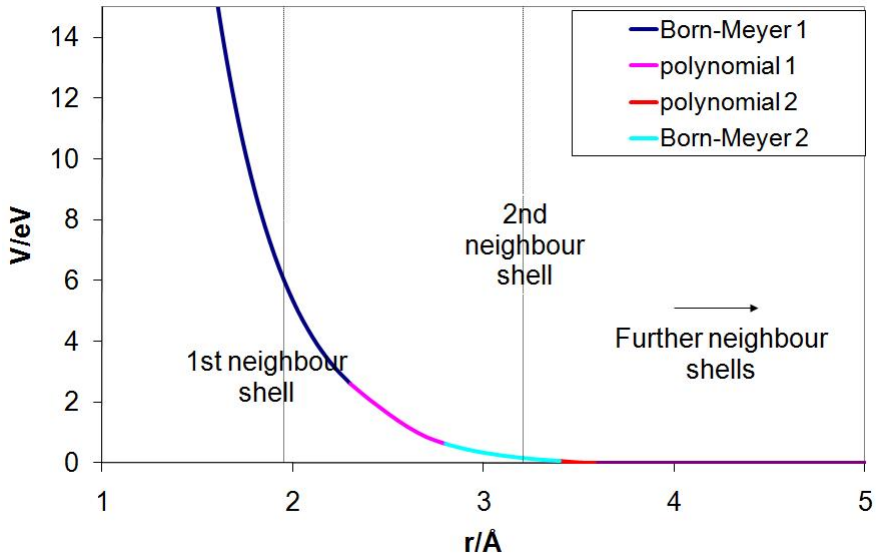


Figure 34: The Ga - N short range potential.

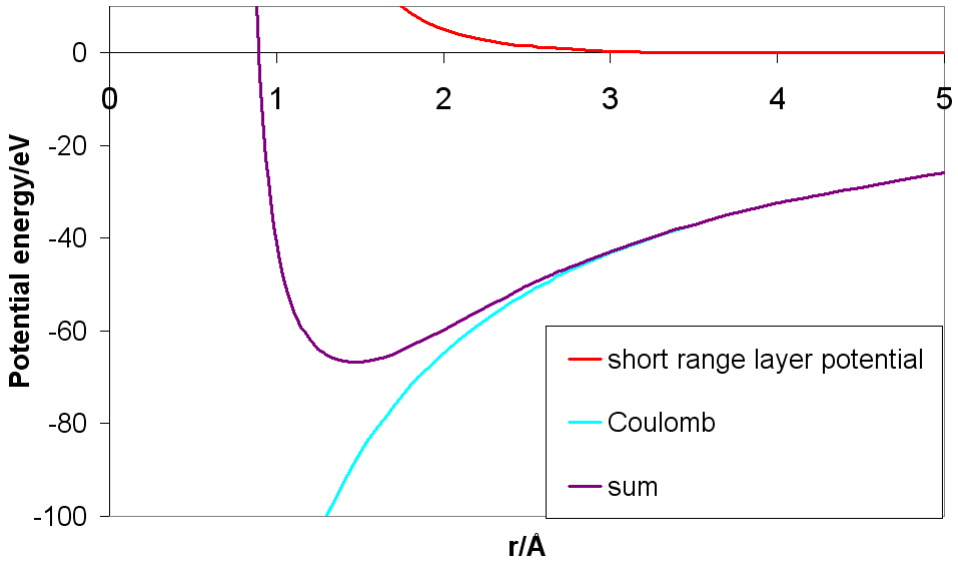


Figure 35: The Ga - N total potential and the short range and long range contributions.

The cation - N potentials are compared in Fig. 36. There is a clear trend of the value of the potential at each interatomic distance with In being the most

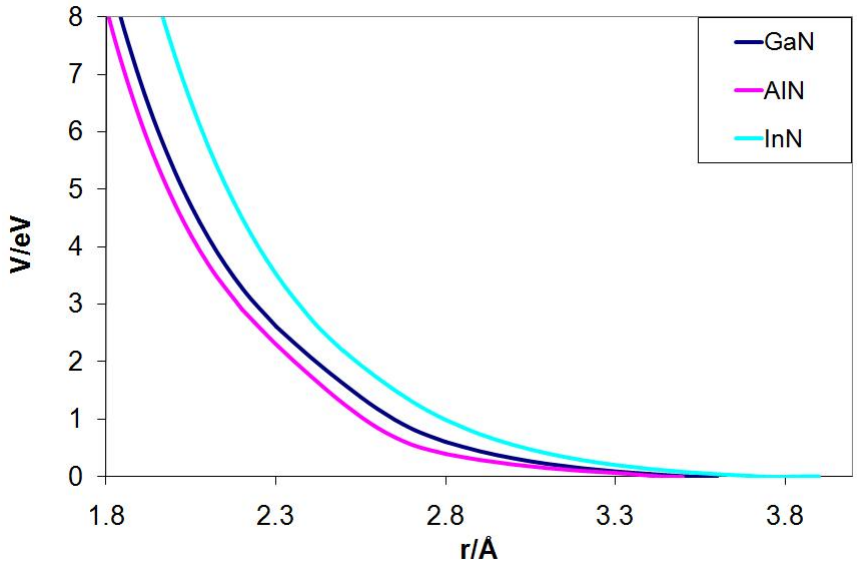


Figure 36: A comparison of cation - N short range potentials.

repulsive and Al the least, which is related to the different lattice parameters and hence bond lengths in the materials.

5.1.2 Results

The results of calculations of bulk physical properties are shown in Tables 2, 3 and 4 on pp 130, 131 and 132. The structural parameters are well predicted, with the possible exception of the AlN and InN a parameters which have a slightly larger than the target 1 % error. The dielectric constants are within approximately 10 % of experimental values.

The elastic off-diagonal constants proved to be more problematic in comparison with experimental data. However, the experimental determination of elastic constants suffers from significant problems as shown by the large scatter in the values reported. Measurements can have large discrepancies as different kinds of samples used (single crystals vs. polycrystalline vs. powder, bulk vs. thin films) with different amounts of defect as well as different techniques (e.g. high/low frequency). The present computational method

Table 2: Comparison of properties of GaN calculated here with other experimental and theoretical values. Values in bold were fitted to.

Property	Present	Experiment	Other calc.
Wurtzite			
latt. const. a (Å)	3.188	**3.19**[1]	3.23[4], 3.20[5],3.18[6]
latt. const. c (Å)	5.186	**5.19**[1]	5.16[4],5.14[5]
u parameter	0.375	**0.377**[2]	0.385[4], 0.375[6]
C_{11} (GPa)	376.1	296±18[7], **390**±15[8], 374[9]	386[4], 410[5], 347[6]
C_{12}(GPa)	238.4	130±10[7], **145**±20[8], 106[9]	160[4], 161[5], 154[6]
C_{13} (GPa)	227.8	158±5[7], **106**±20[8], 70[9]	141[4],142[5], 123[6]
C_{33} (GPa)	496.2	267±17[7], **398**±20[8], 379[9]	391[4],413[5], 381[6]
C_{44} (GPa)	71.9	24±2[7],205±10[8], **101**[9]	115[4],123[5], 81[6]
C_{66} (GPa)	68.8	113[3], **125**[5], 98[6]	
ϵ_{11}^0	9.90	**9.38** [1]	8.05[4], 8.9[10], 8.64[5]
ϵ_{33}^0	10.1	**10.2**[1]	11.2[4]12.64[5]
ϵ_{11}^∞	5.32	**5.35**[10]	5.21[4], 4.25[5]
ϵ_{33}^∞	5.41	**5.35**[1]	5.84[4],4.58[5]
B_0 (GPa)	289	195, 237[2],195[7],210[8],180[9]	210±10[5], 236[5], 208[6]
piezoel.const. e_{33} (Cm^{-2})	5.97	0.65 [12]	0.73 [11]
piezoel.const. e_{31} (Cm^{-2})	-7.14	-0.33 [12]	-0.49 [11]
energy per cation-anion pair/eV	-86.69		
Zincblende			
latt. const. a (Å)	4.48	**4.49**[4]	4.53[4], 4.498[6]
C_{11} (GPa)	304.8	381[3], **296**[4]	300[4],287[6]
C_{12} (GPa)	286.1	130[3], **154**[4]	190[4],169[6]
C_{44} (GPa)	116.8	96[3], **206**[4]	160[4],244[6]
ϵ_{11}^0	10.1	9.7[5]	8.88[4]
ϵ_{11}^∞	5.43	5.3[5]	5.41[4]
B_0 (GPa)	292.4		208[6]
energy per cation-anion pair/eV	-86.60		
Rocksalt			
energy per cation-anion pair/eV	-85.85		

[1] [4] and references therein; [2] [5] and references therein; [3] [6] and references therein; [4] [3] and references therein; [5] [7]; [6] [8]; [7] Sheleg and Savastenko (temperature-dependent broadening of powder X-ray diffraction spectra); [8] Polian (Brillouin scattering); [9] Takagi (Brillouin scattering); [10] Bougrov (2001), 300K; [11] [10]; [12] [30]

Table 3: Comparison of properties of InN calculated here with other experimental and theoretical values. Values in bold were fitted to.

Property	Present	Experiment	Other calc.
Wurtzite			
lattice const., a (Å)	3.577	**3.54**[1]	3.53[1]
lattice const., c (Å)	5.675	**5.69**[1]	5.69[1]
special position, u	0.380	**0.378**[2]	
C_{11} (GPa)	218.6	**190**$\pm$7[5]	271[3], 223[3], 298[1]
C_{12} (GPa)	177.1	**104**$\pm$3[5]	124[3], 115[3], 107[1]
C_{13} (GPa)	143.9	**121**$\pm$7[5]	94[3], 92[3], 109[1]
C_{33} (GPa)	282.1	**182**$\pm$6[5]	200[3], 224[3], 251[1]
C_{44} (GPa)	58.0	10$\pm$1[5]	46[3], 48[3],89[1]
C_{66} (GPa)	20.7		95[1]
ϵ_{11}^0	13.9	**15.0**[1], 15.3[7], 14.4[8]	9.82[1]
ϵ_{33}^0	12.6	13.1[8]	17.71[1]
ϵ_{11}^∞	7.60	**8.4**[1], 9.3[6]	3.5[1]
ϵ_{33}^∞	5.14		3.69[1]
B_0 (GPa)	182.7	126,139[5],139[3]	147[3],141[3]
piezoel.const. e_{33} (Cm^{-2})	6.97	0.97 [9]	
piezoel.const. e_{31} (Cm^{-2})	-3.19	-0.57 [9]	
energy per cation-anion pair/eV	-79.34		
Zincblende			
lattice const., a (Å)	4.99	**4.98**[4]	
C_{11} (GPa)	195.8		242[3]
C_{12} (GPa)	173.3		107[3]
C_{44} (GPa)	0.16		49[3]
ϵ_{11}^0	18.9		
ϵ_{11}^∞	11.7		
B_0 (GPa)	180.8		
energy per cation-anion pair/eV	-79.19		
Rocksalt			
energy per cation-anion pair/eV	-78.87		

[1] [4] and references therein, [2] [5] and references therein, [3] [6] and references therein, [4] [9] and references therein, [5] Sheleg and Savastenko (temperature dependent broadening of powder x-ray diffraction spectra), [6] Tyagai (1977), heavily doped film, [7] Zubrilov (2001), [8] Davydov (1999); [9] [32]

Table 4: Comparison of properties of AlN calculated here with other experimental and theoretical values. Values in bold were fitted to.

Property	Present	Experiment	Other calc.
Wurtzite			
lattice const., a (Å)	3.148	**3.11**[1]	3.11[1]
lattice const., c (Å)	4.976	**4.97**[1]	4.98[1]
special position, u	0.385	**0.382**[2]	
C_{11} (GPa)	444.7	345[5], **411**±10[6]	464[3]398[3],396[3], 417[1]
C_{12} (GPa)	304.7	125[5], 149±10[6]	149[3],140[3],137[3],178[1]
C_{13} (GPa)	306.1	120[5], 99±4[6]	116[3], 127[3], 108[3], 152[1]
C_{33} (GPa)	439.2	395[5], **389**±10[6]	409[3],382[3],373[3], 432[1]
C_{44} (GPa)	102.9	118[5], **125**±5[6]	128[3], 96[3], 116[3]125[1]
C_{66} (GPa)	69.5		120[1]
ϵ_{11}^0	7.88	**8.8**[1], 9.14[7], 8.5[8]	8.07[1]
ϵ_{33}^0	8.86	11.22[1]	
ϵ_{11}^∞	5.21	**4.7**[1], 4.84[7],4.6[8]	4.46[1]
ϵ_{33}^∞	5.38		4.85[1]
piezoel.const. e_{33} (Cm^{-2})	24.6	1.55 [9]	
piezoel.const. e_{31} (Cm^{-2})	-10.7	-0.58 [9]	
B_0 (GPa)	351.6	201[5], 210[6]	228[3], 218[3], 207[3],248[1]
energy per cation-anion pair/eV	-89.31		
Zincblende			
lattice const., a (Å)	4.389	**4.37**[4]	
C_{11} (GPa)	363.3		425[3]
C_{12} (GPa)	344.3		120[3]
C_{44} (GPa)	125.4		112[3]
ϵ_{11}^0	8.28		
ϵ_{11}^∞	5.43		
B_0 (GPa)	350.6	202[4]	
energy per cation-anion pair/eV	-89.19		
Rocksalt			
energy per cation-anion pair/eV	-88.72		

[1] [4] and references therein, [2] [5] and references therein, [3] [6] and references therein, [4] [9] and references therein, [5] Tsubouchi (surface acoustic wave), [6] McNeil (Brillouin scattering), [7] Collins (1967) 300K, reflectivity, [8] Goldberg (2001) 300K; [9] [31]

models high frequency "intrinsic" elastic behaviour. The off-diagonal terms might be improved by the inclusion of three-body terms, which are related to a degree of covalency in the system and take account of bond rotations. Importantly, the correct relative stability of the three commonly studied phases is reproduced, the accuracy of which can be improved by exploiting the fact that the different phases have varying numbers of neighbours at particular interatomic distances as can be seen from Tables 42, 43 and 44. In particular, the wurtzite structure has one neighbour at a cation-N distance which is not found in the other two structures. In wurtzite GaN, a Ga-N pair is seen at an interatomic distance of about 3.2Å; there are no such pairs in zincblende and rocksalt. Shifting the value of the potential by a constant offset at this interatomic distance[10] corrects the relative phase stability if necessary and can be used to control the pressure at which transitions between the phases occur. One must however be careful not to introduce unphysical local maxima into the potential form by introducing such offsets.

The dependence of the enthalpy per cation-anion pair on pressure for GaN phases is shown in Fig. 37. Our calculated GaN phase transition from wurtzite to rocksalt is 38.4 GPa, in good agreement with the lower end of experimental data of 37 GPa from [11]. A higher value of about 52.2 GPa has been reported in [12]. The calculated AlN transition pressure is 24.6GPa; X-ray diffraction experiments reported in [13] measure the transition at 22.9 GPa. The InN calculated pressure is 15.1GPa, compared with the experimental value of 12.1 GPa from [12]. The results are summarised in Table 5.

The introduction of offsets as a way of improving the fit to transition pressures is an attractive option but this correction needs to be interpreted carefully. Seemingly accurate transition pressures may mask problems with the potentials. In the present case the agreement with a wide range of experimental properties is good and the potentials have physically sensible functional forms so one can be satisfied with the quality of the model.

The dependence of volume on pressure, i.e. the equation of state (EoS), is explicitly shown in Fig. 38 on p. 135. The GaN experimental EoS is taken

[10] i.e. moving the relevant section of a curve such as the one in Fig. 34 up or down in energy, and adapting the adjacent polynomial region accordingly.

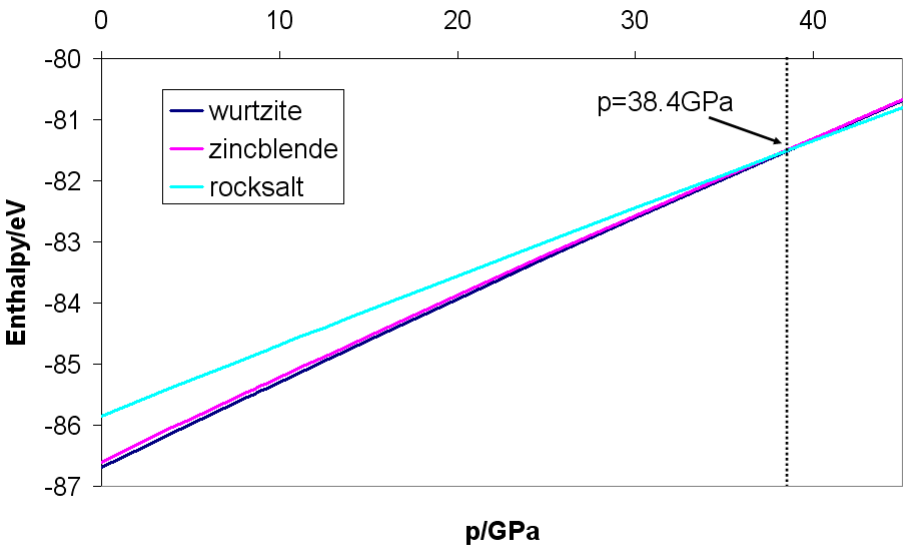


Figure 37: The relative stability of GaN phases. The diagram indicates a phase transition from wurtzite to rocksalt at 38.4GPa. The corresponding transition pressure for AlN is 24.6GPa and for InN 15.7GPa

Table 5: Wurtzite-rocksalt transition pressure

	Transition pressure (GPa)	
Compound	Present	Experiment
AlN	24.6	24.6 [13]
GaN	38.4	37 [11], 52.2 [12]
InN	15.1	12.1 [12]

from [14] where the data is fitted to the third-order Birch-Murnaghan equation of state

$$P(V) = \frac{3}{2}\left[\zeta^{7/3} - \zeta^{5/3}\right]\left[1 + \frac{3}{4}(B_0' - 4)(\zeta^{2/3} - 1)\right] \tag{85}$$

where V_0 is the volume at ambient pressure, $\zeta = (V_0/V)$ and $B_0' = dB_0/dP$. Reasonable agreement with experimental data is shown. As both the a and the c parameter contribute to the volume, the compressibility along both directions will shed some light on the deviations from experimental data in

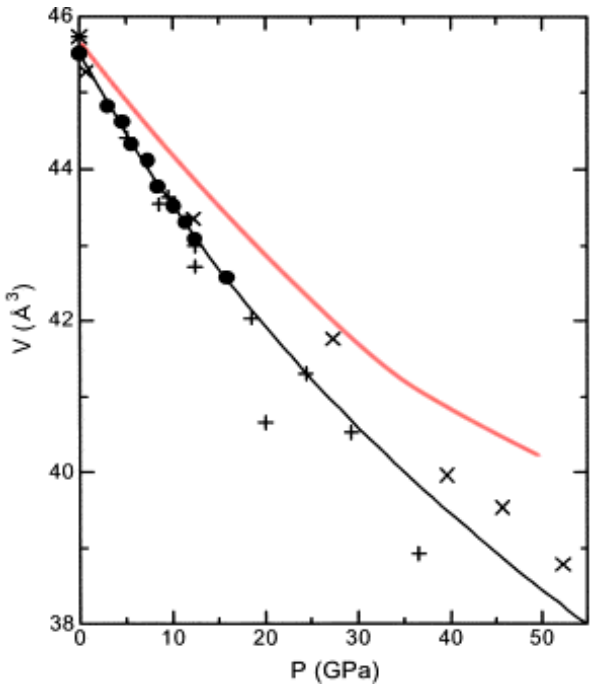


Figure 38: The equation of state for wurtzite GaN calculated with our model (red) and experimental data (black circles and crosses) in [14]. The black line is a result of fitting the experimental data to the Birch-Murnaghan equation of state.

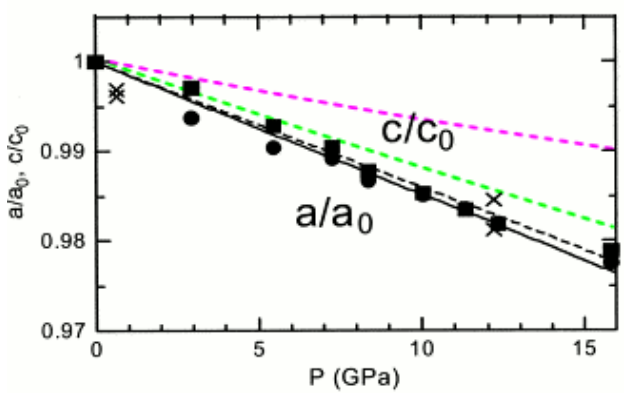


Figure 39: The pressure dependence of the compressibilities for GaN. The present calculation is in colour (green for a parameter and pink for c parameter), and the results in black are from [14]. a_0 and c_0 are the values of the cell parameters at zero pressure.

the Equation of State calculation. The compressibilities are discussed next. The volume in Fig. 38 is a function of the a and c parameters. Their pressure dependences and respective compressibilities are shown in Fig. 39. The a parameter pressure dependence is very well reproduced. Larger errors occur in the calculation of the c parameter under pressure, which can be traced to the poor reproduction of the c_{33} elastic constant. It has proven difficult to fit some of the elastic constants while at the same time ensuring compatibility with the QM calculation (see Chapter 8). These problems, described in more detail in Section 8.2.3 on p. 201, are partly related to the highly manual process which is at present required to create a potential model which both fits experimental data and does not suffer from distortions in QM/MM calculations. Automation of the process might result in an improved parameter set. Γ point phonon frequencies calculated with the present model are compared with experimental data and other calculations in Tables 6, 7 and 8. Like the elastic constants, the phonons are a measure of the second derivatives of energy. The present results are in good agreement with experiment usually within 10%, with the exception of the E_2(low) and B_1(low) modes which have more significant errors.

The dispersion curves and density of state are shown in Fig. 40, 41 and 42. The general features of the dispersion curves are correct but they are not reliable on the quantitative level. Specifically, GaN and InN experimental results have a large gap between the low-energy and the high-energy modes which is underestimated with the present model.

Table 6: Comparison of phonon frequencies in cm^{-1} at Γ point for AlN with experimental data and other calculations

Mode	Present	Experiment	Other calc.
Wurtzite			
E_2(low)	192.3(doublet)	248.4[1], 303[2] 241[2], 252[2]	236[2], 228[2], 301[2]
B_1(low)	595.8		553[2], 524[2] 723[2], 580[1]
A_1(TO)	670.5	659[2], 667[2], 660[2] 607[2], 614[2]	629[2], 601[2], 668[2]
E_1(TO)	712.7	673.4[1],614[2], 671[2] 667[2], 672[2], 673[2]	649[2], 650[2], 734[2]
E_2(high)	742.2(doublet)	660[1], 426[2],665[2] 660[2], 660[2]	631[2], 638[2], 704[2]
B_1(high)	801.9		717[2], 703[2], 772[2]
A_1(LO)	855.6		
E_1(LO)	874.8	917[1]	
Zincblende			
TO	663.4		551[2], 558[2], 603[2], 600[2]

[1] [15] and references therein; experiments at $6K$, [2] [16] and references therein

Table 7: Comparison of phonon frequencies in cm^{-1} at Γ point for GaN with experimental data and other calculations

Mode	Present	Experiment	Other calc.
Wurtzite			
E_2(low)	108.8 (doublet)	143.6[1]	153[1], 143[1]
B_1(low)	432.9	319[1], 337[1]	
A_1(TO)	556.2	546[1], 533.8[1]	546[1], 541[1]
E_1(TO)	546.6	560.2[1]	554[1],568[1]
E_2(high)	577.2(doublet)	569[1]	565[1],579[1]
B_1(high)	674.1	728[1], 720[1]	
A_1(LO)	761.4	736[1]	734[1], 748[1]
E_1(LO)	746.1	744[1]	739[1], 757[1]
Zincblende			
TO	552.4	551[2], 558[2], 603[2], 600[2]	

[1] [15] and references therein; experiments at $6K$, [2] [16] and references therein

Table 8: Comparison of phonon frequencies in cm^{-1} at Γ point for InN with experimental data and other calculations

Mode	Present	Experiment	Other calc.
Wurtzite			
E_2(low)	69.4(doublet)	87[1]	
B_1(low)	274.8		
A_1(TO)	402.3	447[1]	
E_1(TO)	421.4	476[1]	
E_2(high)	451.4(doublet)	488[1]	
B_1(high)	571.0		
A_1(LO)	627.6	586[1]	
E_1(LO)	570.0	593[1]	

[1] see [17]

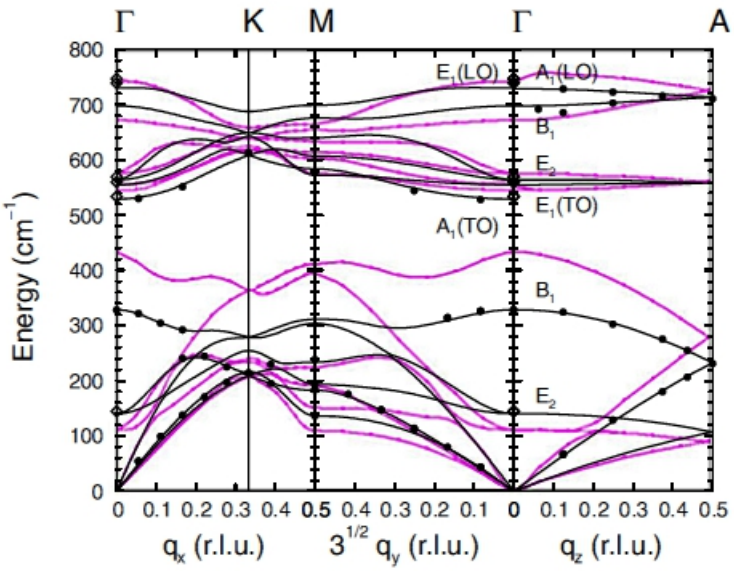


Figure 40: The calculated and experimental GaN phonon dispersion curves. The pink curve is calculated with the present model and the black curves are experimental and *ab initio* data taken from [18].

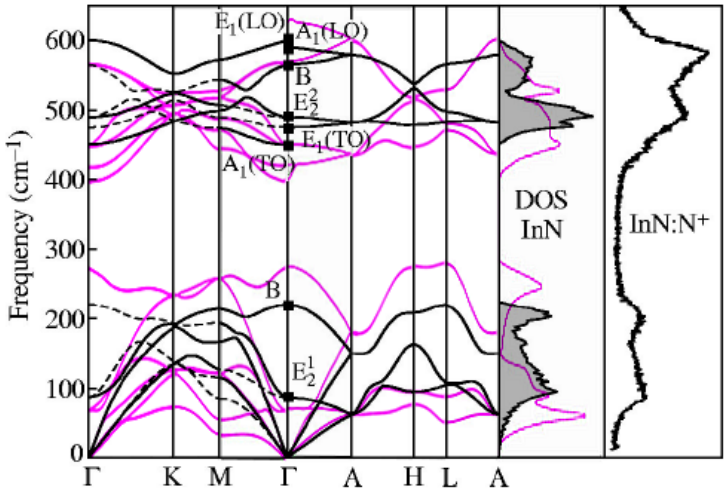


Figure 41: The calculated and experimental InN phonon dispersion curves. The pink curve is calculated with the present model and the black curves are experimental and *ab initio* data taken from [17] .

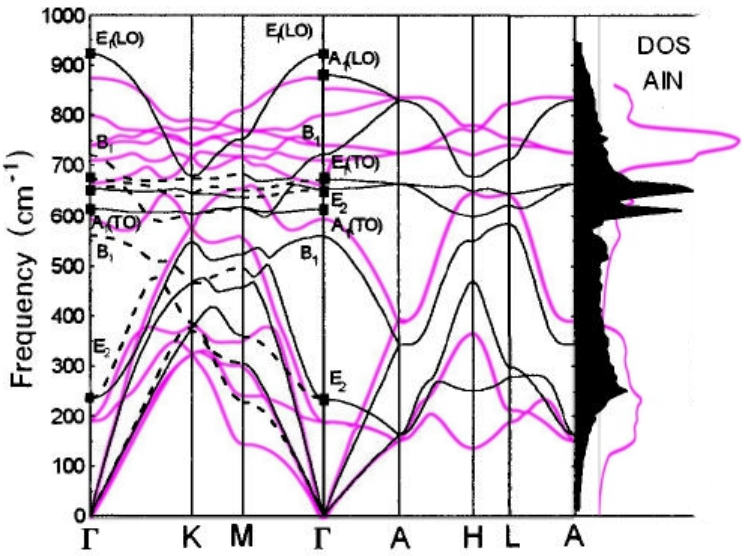


Figure 42: The calculated and experimental AlN phonon dispersion curves. The pink curve is calculated with the present model and the black curves are experimental and *ab initio* data taken from [15].

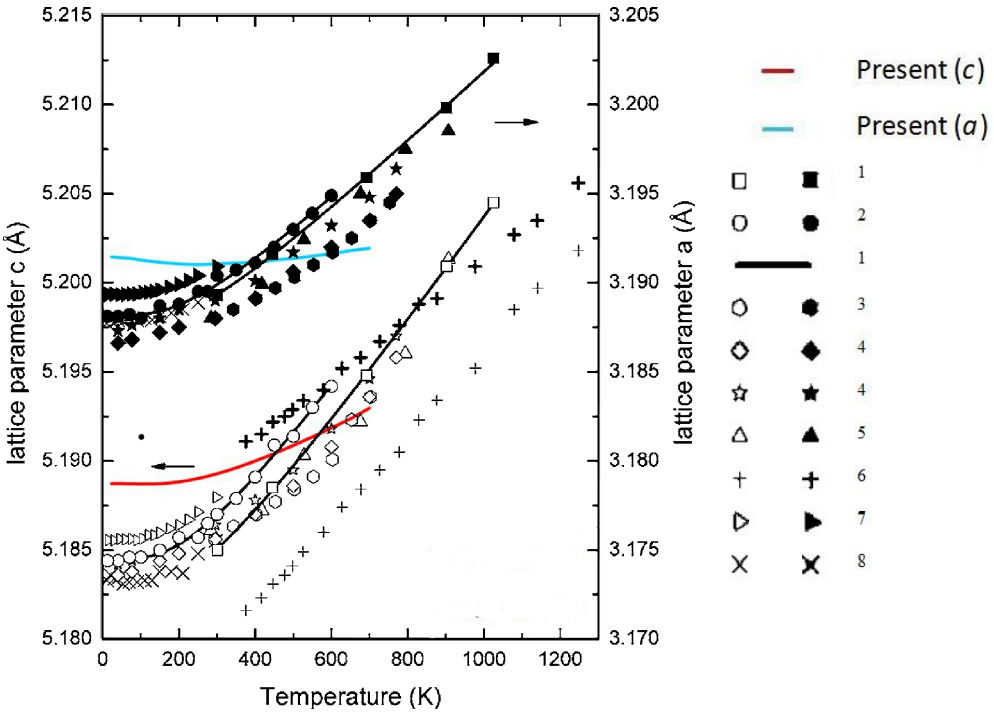


Figure 43: The structure parameters as a function of temperature.

[1] [19] ,[2] [20] ,[3] [21],[4] [22],[5] [23] ,[6] [24], [7] [25], [8] [26]

The errors in the phonon dispersion translate into errors in temperature behaviour in Fig. 43 and 44,which can be seen from the temperature dependences of the structural parameters and the thermal expansion coefficients α

$$\alpha = \frac{1}{x}\frac{\partial x}{\partial T} \tag{86}$$

where T is the temperature and x is one of the structural parameters. In the present work α was calculated simply as

$$\alpha = \frac{1}{x_{T+\delta T}}\frac{x_{T+\delta T} - x_T}{\delta T} \tag{87}$$

The thermal expansion coefficients are of the correct order of magnitude but clearly underestimated. Modelling correct temperature behaviour with interatomic potentials that also simulate other properties well is, however, a

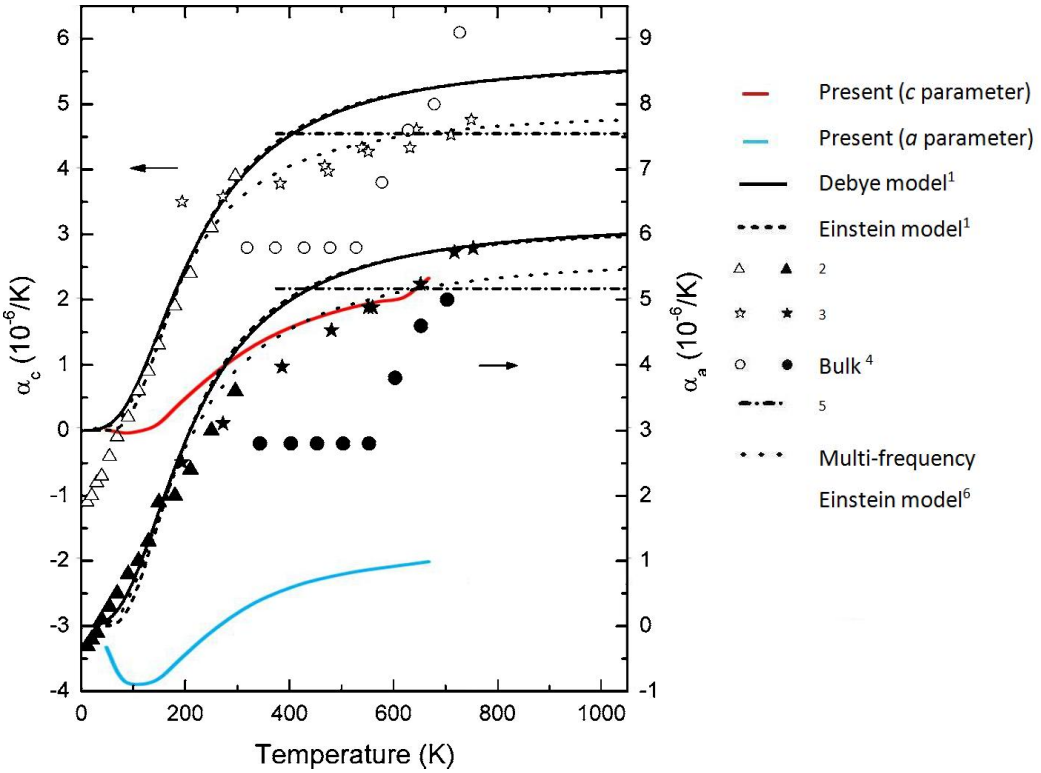


Figure 44: The thermal expansion coefficients as a function of temperature. [1] [19] ,[2] [26] ,[3] [29],[4] [21],[5] [24] ,[6] [25]

very difficult task and not essential for the purposes for which our models will be used.

5.1.3 Summary

In this section, new interatomic potentials for modelling GaN, AlN and InN, were presented and extensively tested in the calculation of a wide range of physical properties. The models show reasonable agreement with experimental structural and dielectric properties as well as phonon frequencies. Properties such as phonon densities of states and temperature behaviour can be described only with limited accuracy. The off-diagonal and c_{33} elastic constants also show less good agreement, leading to errors in the compressibility

along the c-axis. A possible route to further improvement of the potential would be to include three body terms as opposed to just two body interactions as in the current model. Furthermore, an automated process of fitting potentials which are suitable for use in QM/MM applications, i.e. which do not lead to boundary distortions, is suggested.
In future work the study of the relative stability of phases may be extended to other polytypes of the material. Polytypes are common in materials such as silicon carbide and micas. [27] discusses the 4H phase in semiconductors while [28] analyses the 9R phase in epitaxial layers, which might have potentially stabilising effects on phases which are metastable in the absence of strain.

References

[1] GULP: A computer program for the symmetry-adapted simulation of solids, J.D.Gale, J.Chem.Soc.,Faraday Trans. 93 **629 - 637** (1997)

[2] General Utility Lattice Program, J.D.Gale `https://www.ivec.org/gulp/help/gulp3.4_manual.pdf`

[3] An interatomic potential study of the properties of gallium nitride, P.Zapol, R.Pandey, J.D.Gale, J.Phys.: Condens. Matter 9 **9517 - 9525** (1997)

[4] Classical simulations of the properties of group-III nitrides, J.A. Chisholm, D.W.Lewis, P.D.Bristowe, J.Phys.Condens.Matter 11 **L235 - L239** (1999)

[5] Ab initio phonon dispersions of wurtzite AlN, GaN and InN, C.Bungaro, K.Rapcewicz, J.Bernholc, Phys. Rev. B 61 **6720 - 6725** (2000)

[6] Elastic properties of zinc-blende and wurtzite AlN, GaN, and InN, A.F.Wright, J.Appl.Phys.82 **2833 - 2839** (1997)

[7] Properties of advanced semiconductor materials: GaN, AlN, InN, BN, SiC, SiGe, M. E.Levinshtein, S. L. Rumyantsev, M.Shur (2001)

[8] Modelling compound semiconductors:analytical bond-order potential for gallium, nitrogen and gallium nitride, J.Nord, K.Albe, P.Erhart, K.Nordlund, J.PHys.Condensed Matter 15 **5649 - 5662** (2003)

[9] Elastic Properties of Zinc-blende GaN, AlN and InN from Molecular Dynamics, F. Benkabou, M. Certier, H. Aourag, Molecular Simulation 29 **201 - 209** (2003)

[10] Spontaneous polarization and piezoelectric constants of III-V nitrides, F.Bernardini, V.Fiorentini, D.Vanderbilt, Phys.Rev.B 56 **10024 - 10027** (1997)

[11] Pressure induced rocksalt phase of aluminum nitride: A metastable structure at ambient condition, Q.Xia, H.Xia, A. L. Ruoff, J. Appl. Phys. 73 **8198 - 8200** (1993)

[12] Stability of the wurtzite-type structure under high pressure: GaN and InN, M.Ueno, M. Yoshida, A.Onodera, O.Shimomura, K.Takemura, Phys. Rev. B 49 **14 - 21** (1994)

[13] X-ray observation of the structural phase transition of aluminum nitride under high pressure, M. Ueno, A. Onodera, O. Shimomura, K. Takemura, Phys. Rev. B 45 **10123 - 10126** (1992)

[14] Precise measurement of equation-of-state and elastic propertoes for GaN up to 16GPa, T.tsuchiya, K.Kawamura, O.Ohtaka, H.Fukui, T.Kikegawa, Sol.St.Comm. 121 **9 - 10** (2002)

[15] Phonon dispersion and Raman scattering in hexagonal GaN and AlN, V.Yu.Davydov, Yu.E.Kitaev, I.N.Goncharuk, A.N.Smirnov, J.Graul, O.Semchinova, D.Uffmann, M.B.Smirnov, A.P.Mirgorodsky, R.A.Evarestov, Phys.Rev.B58 **12899 - 12907** (1998)

[16] Optical phonon modes in GaN and AlN, I.Gorczyca, N.E.Christensen, E.L.Peltzer y Blanca, C.O.Rodriguez, Phys.Rev,B 51 **11936 - 11939** (1995)

[17] Experimental and theoretical studies of phonons in hexagonal InN, V.Yu.Davydov, V.V.Emtsev, I.N.Goncharuk, A.N.Smirnov, V.D.Petrikov, V.V.Mamutin, V.A.Vekshin, S.V.Ivanov, M.B.Smirnov, T.Inushima, Appl.Phys.Letters 75 **3297 - 3299** (1999)

[18] Phonon dispersion curves in wurtzite structure GaN determined by inelastic x-ray scattering, T.Ruf, J.Serrano, M.Cardona, P.Pavone, M.Pabst, M.Krisch, M.D'Astuto, T.Suski, I.Grzegory, M.Leszczynski, Phys.Rev.Lett. 86 **906 - 909** (2001)

[19] Temperature dependence of the thermal expansion of GaN, C. Roder, S. Einfeldt, S. Figge, D. Hommel, PRB 72 **085218** (2005)

[20] Thermal expansion of bulk and homoepitaxial GaN, V.Kirchner, H. Heinke, D.Hommel, J.Z.Domagala, M.Leszczynski, APL 77 **1434 - 1436** (2000)

[21] Thermal expansion of gallium nitride, M. Leszczynski, T. Suski, H. Teisseyre, P. Perlin, I. Grzegory, J.Jun, S. Porowski, and T. D. Moustakas, J. Appl. Phys. **4909 - 4911** 76 (1994)

[22] Thermal Expansion of GaN Bulk Crystals and Homoepitaxial Layers, M. Leszczynski, H. Teisseyre, T. Suski, I. Grzegory, M. Bockowski, J. Jun, B. Palosz, S. Porowski, K. Pakula, J. M. Baranowski, A. Barski, Acta Phys. Pol. A 90 **887 - 890** (1996)

[23] The preparation and properties of vapor-deposited single crystalline GaN, H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. 15 **327** (1969)

[24] Thermal expansion of GaN, E.Ejder, physica status solidi (a) 23 **K87 - K90** (1974)

[25] Lattice parameters and thermal expansion of GaN, R.R. Reeber, K.Wang, J.Mat.Res. 15 **40 - 44** (2000)

[26] W. Paszkowicz, J. Z. Domagala, J. A. Sokolowski, G. Kamler, S. Podsiadlo, and M. Knapp, in Synchrotron Radiation Studies of Materials: Proceedings of 5th National Symposium of Synchrotron Radiation Users, edited by M. Lefeld-Sosnowska and J.Gronkowski Zaklad Graficzny Uniwersytefu Warszawskiego, Warsaw, 1999, pp. 183189.

[27] Monte Carlo modelling of wurtzite and 4H phase semiconducting materials, K.F.Brennan, E.Bellotti, M.Farahmand, H.-E.Nilsson, P.P.Ruden, Y.Zhang, VLSI design 13 **117 - 114** (2001)

[28] Polytypism in epitaxially grown gallium nitride, H.Selke, V.Kirchner, H.Heinke, S.Einfeldt, P.L.Ryder, D.Hommel, J.Cryst.Growth **57 - 64** (2000)

[29] A. U. Sheleg and V. A. Savastenko, Vestsi Akad. Navuk BSSR,Ser. Fiz.-Mat. Navuk 3, 126 **1598** (1976)

[30] Elastic strain relaxation and piezoeffect in GaN-AlN, GaN-AlGaN and GaN-InGaN superlattices, A.D.Bykhovski, B.L. Gelmont, M.S. Shur, J. Appl. Phys. 81 **6332-6338** (1997)

[31] On the properties of AlN thin films grown by low temperature reactive rf sputtering L.Xinjiao, Xu Zechuan, He Ziyou, Cao H'uazhe, Su Wuda, Chen Zhongcai, Zhou Feng, Wang Enguang, Thin Solid Films 139, **261-274** (1986)

[32] Spontaneous polarization and piezoelectric constants of III-V nitrides, F. Bernardini, V. Fiorentini, D.Vanderbilt, Phys. Rev. B 56 **R10024-R10027** (1997)

6 Point defects

Defects in III-V nitride alloys are of high importance as they impart useful as well as deleterious properties to the devices in which these materials are used. They are the basis of ionic conductivity and can increase as well as decrease carrier concentrations. Defects may also provide a mechanism for reducing strain in the material.
Defect calculations also enable us to test the potentials for robustness since the interatomic distances in defect systems can be quite different from the equilibrium distances for which the potentials were parameterised. Furthermore, the reduction of symmetry in a defect calculation may reveal instabilities that are not apparent in the bulk.
In this section the interatomic potentials from the previous section will be used in conjunction with the Mott-Littleton method to calculate point defect formation energies and geometries in GaN in the formal charge state. Calculations of other charge states are inaccessible to these potentials as parameterising the interactions of a non-formal-charge defect with the rest of the lattice would be very difficult. Cation substitutional defect solution energies will also be evaluated.

6.1 Methodology

Defect energies in GULP are calculated using the Mott-Littleton method (see Section 2.1.2) as implemented in the GULP package. The harmonic approximation for region 2 displacements of the Mott-Littleton method only holds if region 1 is large enough. Energy convergence checks were performed to ensure that the defect energies were converged to less than 0.1eV. The required region 1 radii were approximately 21 Å.
A further methodological issue is the initial defect geometry. If this is too different from the geometry associated with the global minimum, the structure may optimise to a local minimum. A carefully selected range of starting geometries was explored in the search for a global minimum. The nitrogen interstitials had starting coordinates taken from a mesh of $4 \times 4 \times 4$ equally

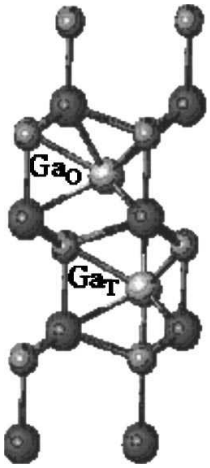


Figure 45: Positions of Ga octahedral and tetrahedral interstitial defects in wurtzite structure. Analogous N interstitials can be formed although in calculations they are predicted to assume a split interstitial position. From [1].

spaced fractional coordinate points in the unit cell (such that $x \geq y$ due to the ab-plane symmetry of the wurtzite unit cell. The configuration with the lowest resulting energy was taken. Cation interstitials in previous work have been found to occupy an octahedral site which lies along the hexagonal channel in the c-direction (see Fig. 45). The channel was explored for the lowest energy configurations in addition to performing the unit cell search described above.

The defect energy calculated in GULP is the difference between the energy of the system with defect U^d and the perfect system U^p , i.e.

$$U_{defect} = U^d - U^p \tag{88}$$

For an interstitial, this corresponds to the energy change on taking an ion at infinity to the equilibrium position in the lattice. This is, however, not a quantity directly measurable by experiment. Comparison with experimental values will require other terms usually via a Born Haber cycle. Some of the energies in such cycles, such as ionisation and atomisation energies, cannot be extracted from the current model. However, combinations of defect energies, in particular, Schottky and Frenkel energies can be measured experimentally

and calculated with the current model.
The Schottky defect energy, E_{Schottky}, is defined as

$$E_{\text{Schottky}} = E_{\text{cation vacancy}} + E_{\text{anion vacancy}} + E_{\text{lattice}} \tag{89}$$

where $E_{\text{cation vacancy}}$ and $E_{\text{anion vacancy}}$ are the cation and anion defect energies and E_{pair} is the exothermic lattice energy per anion-cation pair.
The cation Frenkel energy, E_{Frenkel}, can be expressed as

$$E_{\text{Frenkel}} = E_{\text{cation vacancy}} + E_{\text{cation interstitial}} \tag{90}$$

and analogously for the anion.
To calculate the solution energies we used the supercell approach with $4 \times 4 \times 4$ supercells, containing 256 atoms. One of the cations was replaced with another element; for example for GaN lattice, one Al or one In substitutional atom was placed at a Ga site. The solution energy E_{solution} is then calculated as

$$E_{\text{solution}} = E_{\text{supercell}} - 127E_{\text{GaN}} - E_{\text{lattice impurity}} \tag{91}$$

where $E_{\text{supercell}}$ is the energy of the supercell as calculated by GULP. E_{GaN} is the lattice energy per cation-anion pair in a pure binary compound (without a substitutional impurity) and $E_{\text{lattice impurity}}$ is the lattice energy of the impurity binary compound. With increasing supercell size the infinite dilution limit is approached.

6.2 Results and discussion

In Fig. 46, the variation in energy vs. region size for the N interstitial in GaN is presented. At a region 1 size of about 20 Å the desired convergence to 0.1eV is achieved. It is difficult to perform larger calculations as GULP experiences memory problems.

The final geometries of interstitials and vacancies are shown in Table 9. Antisites were not studied in this work due to their large formation energies, as

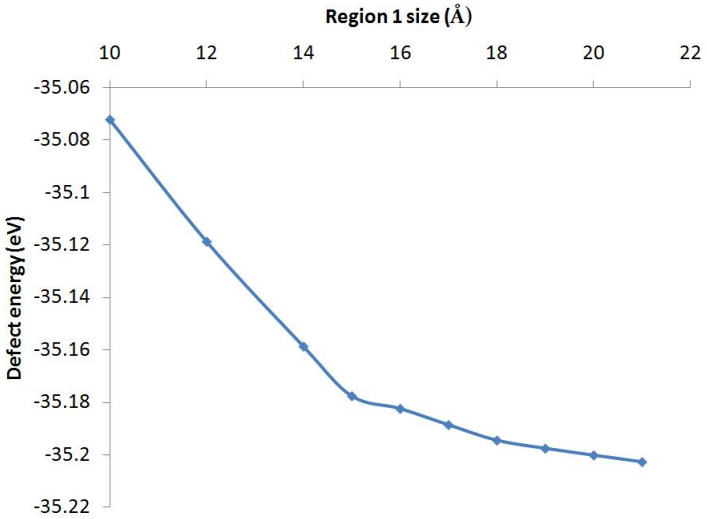


Figure 46: The variation in defect energy vs. region size (convergence curve) for the N interstitial in GaN.

suggested in the literature. The Ga interstitial is found to be lowest in energy at the octahedral side. N interstitial forms a split interstitial with another lattice N. The nearest neighbours of the vacancies experience a breathing relaxation outwards from the defect centre.

Nitrogen interstitials have been widely reported to form a split interstitial configuration, e. g. in [2], as shown in Fig. 47 although [3] discusses a N^{3-} interstitial in a non-split interstitial configuration. The distance between the lattice N and the interstitial is measured and compared with the bond length in the N_2 molecule, 1.11Å, in Table 9. This comparison is frequently used when less negative charge states are studied. In those cases the bond length is comparable to that of a nitrogen molecule. For N^{3-} the Coulomb repulsion between the nitrogens is very strong, hence the bond length is more than twice the N-N distance in the N_2 molecule.

Our calculations do show a split interstitial in that the lattice nitrogen is significantly displaced from its lattice site. The interstitial N is in the centre of the channel as shown in Fig.48 similar to the configuration discussed in [3].

The N-N bond lengths increase in the order Al<Ga<In, reflecting increas-

Table 9: Native defects: geometries and energies of the lowest energy configurations. The split N interstitial bond is compared to the bond length of free N_2 of 1.11 Å. The % change in vacancies measures the breathing relaxation of the distance between the defect centre and the nearest unlike neighbours. Three of these neighbours are equivalent and one is inequivalent, positioned along the z-axis from the defect centre. The cation interstitial bond lengths are compared with the nearest neighbour distances in the bulk. The defect energy in the last column is as calculated in GULP via equation 88.

Defect type	Geometry (Å)	% change	Defect energy (eV)
GaN			
N interstitial	lattice split N 2.50	+125	-35.20
N vacancy	Ga equivalent 2.40 (3)	+23	53.54
	Ga inequivalent 2.39	+22	
Ga interstitial	N 2.01 (3), 2.23 (3)	+3, +15	-29.04
	Ga 2.43 (3) Ga 2.70 (3)	-24, -15	
Ga vacancy	N equivalent 2.21 (3)	+14	48.49
	N inequivalent 2.28	+17	
AlN			
N interstitial	lattice split N 2.43	+119	-36.78
N vacancy	Al equivalent 2.37 (3)	+24	56.42
	Al inequivalent 2.32	+21	
Al interstitial	N 2.13 (3), 2.00 (3)	+12, +5	
	Al 2.37 (3), 2.64 (3)	-23, -14	-29.93
Al vacancy	N equivalent 2.12 (3)	+11	50.67
	N inequivalent 2.19	+15	
InN			
N interstitial	lattice split N 2.77	+150	-33.02
N vacancy	In equivalent 2.61 (3)	+20	46.71
	In inequivalent 2.77	+28	
In interstitial	N 2.10 (3), N 2.72 (3)	-3,+25	
	In 2.82 (3), In 2.91 (3)	-20, -17	-29.34
In vacancy	N equivalent 2.60 (3)	+20	44.36
	N inequivalent 2.82	+30	

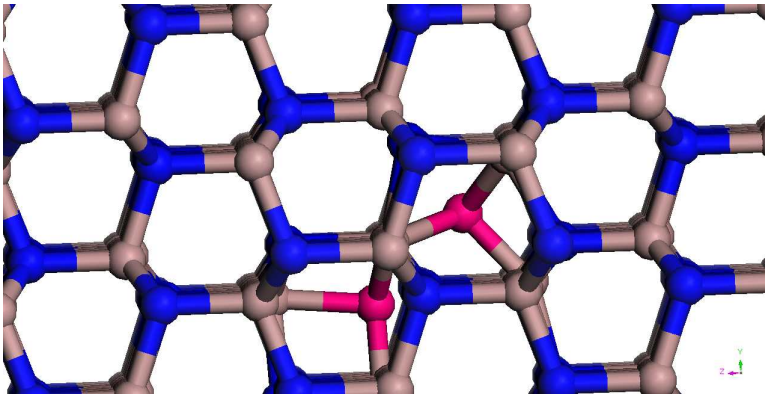


Figure 47: N split interstitial interstitial in GaN (view along the x-axis). The interstitial and the lattice N with which it forms a bond are shown in pink.

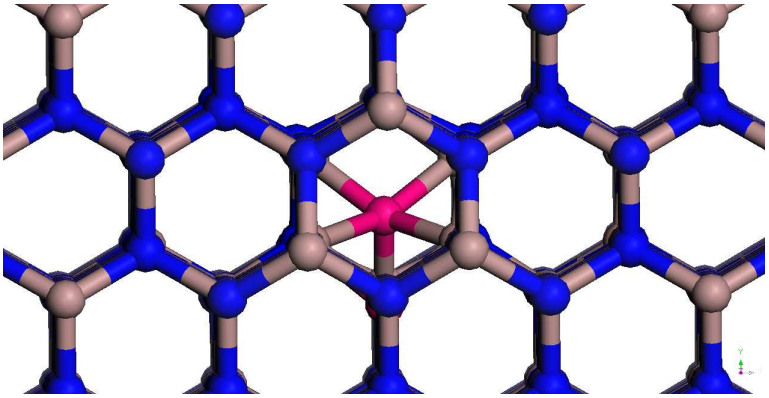


Figure 48: The geometry of the split N interstitial (view along the z-axis). The interstitial and the lattice N with which it forms a bond are shown in pink.

ing lattice parameters. The interatomic distances of the N interstitial and the lattice nitrogens are significantly different from the bulk N-N distances (around 3.1 Å for GaN and AlN and 3.5 Å for InN).

Earlier periodic DFT calculations suggest that N^{3-} interstitial has very high formation energies [2]. Our own QM/MM calculations in Chapter 8 seem to agree that the 3- charge state is unstable (see Table 29).

The cation interstitials can occupy two high symmetry sites, referred to as octahedral and tetrahedral (see Fig. 45). The octahedral arrangement has consistently been reported as the lower energy configuration as has indeed been shown in the present work. There are six nearest N neighbours in this

arrangement, two sets of three N ions by symmetry, and two sets of three nearest cation neighbours, the distances to which are reported in Table 9 and compared with the bulk bond lengths. The cation-cation bond lengths are generally about 20% shorter than in the bulk and cation-N bonds about 10% longer.

In GaN each N vacancy site is surrounded by four nearest Ga neighbours, three in-plane, equivalent ones and one along the c-axis which is inequivalent. The distances of these two types of neighbour from the defect centre are reported in Table 9 and compared with equilibrium bond lengths. For all the defects the relaxation is outwards from the defect centre, which is consistent with the reduction in the Coulomb attraction of the nearest neighbour Ga ions to the defect centre after the N is removed. Similar observations hold for the other vacancies. We also note here that our QM/MM calculations, reported in Chapter 8, indicate that the formally charged N vacancy may be unstable.

As indicated in Table 9, the deviations of the interatomic distances in defects with respect to the bulk bond lengths can be significant. Robust potentials which interpolate well between the bulk interatomic distances are needed if reliable defect geometries are to be predicted. The deviations from experimental data observed in the potentials under pressure in section 5 indicate that there is some room for improvement at non-equilibrium distances, for instance by further fitting the structural parameters to pressure data.

The Schottky and Frenkel defect energies are reported in Table 10. Schottky defect energies are the lowest for all three materials, indicating that vacancies are the dominant type of defect in III-V nitrides, which is consistent with the findings of DFT calculations, which also predict vacancies to be thermodynamically most stable (see Section 4.3 and 8).

The cation Frenkel defects have the highest energies. The same trends are observed in [4], apart from the Frenkel cation and anion defect energies in AlN. A direct comparison of defect energies is not possible as the potential used in [4] is a partial charge model with ionic charges of 2+ and 2-. The Schottky and Frenkel defect energies become larger in the order Al>Ga>In, consistent with the findings in [4]. Our reported values are generally larger

Table 10: Schottky and Frenkel defect energies per atom. The values in brackets are from [4], where partial ionic charges are used as opposed to the formal ionic charges in the present work.

Compound	Schottky (eV)	N Frenkel (eV)	Cation Frenkel (eV)
AlN	8.89 (5.17)	9.82(7.4)	10.37(6.62)
GaN	7.67 (4.74)	9.17 (6.66)	9.72 (7.42)
InN	6.00 (3.39)	6.94(4.76)	7.58 (7.43)

Table 11: Solution energies per cation. The calculations labelled PBC are based on a $4 \times 4 \times 4$ supercell. The M-L label denotes Mott-Littleton defect energies. The values in brackets are from [4], based on 96-atom supercells.

	Al		Ga		In	
Compound	PBC	M-L	PBC	M-L	PBC	M-L
AlN	×	×	0.04(0.86)	1.48	1.07(0.51)	10.78
GaN	0.04 (0.99)	-1.40	×	×	0.62(1.53)	8.93
InN	0.74 (0.09)	-9.17	0.46(1.00)	-7.92	×	×

than those obtained for the partial charge systems.

Finally, the solution energies are shown in Table 11. The corresponding Mott-Littleton energies are also shown. The solution energies are in general less endothermic than the values from [4], based on a partial charge model and 96-atom supercell compared with a 256-atom supercell in our calculation. It should be recalled here that InN has a different spring constant from AlN and GaN in this model, leading to ambiguity in its definition and possible inaccuracies. A possible improvement would involve changing the N spring constant based on the type of nearest cation neighbours.

The endothermic solution energies imply that the III-V nitrides are not soluble in each other even at low concentrations. These calculations are implicitly performed at 0K and do not model the increased solubility of an impurity with temperature. The solubility is explored further in the next chapter, in the context of alloys.

6.3 Summary

In this work the geometries and defect energies of native defects were studied with interatomic potentials from Chapter 5. Schottky defects were found to be energetically most favourable, consistent with DFT predictions of vacancies being more abundant in GaN [5]-[9]. Further discussion and comparison with our QM/MM calculations is given in Chapter 8.
The solution energies of the cations in pure binary compounds were of the order of 1eV.
In the present work diffusion barriers were not studied. Diffusion is an important phenomenon and would be an interesting topic for future exploration. Likewise defect complexes, line defects and dislocations are possible future directions for this work.

References

[1] Intrinsic defect properties in GaN calculated by *ab initio* and empirical potential methods, F.Gao, E.J. Bylaska, W. J. Weber, Phys.Rev.B **245208** (2004)

[2] First-principles calculations for defects and impurities: Applications to III-nitrides, C. G. Van de Walle J. Neugebauer, J.Appl.Phys. 95 **3851 - 3879** (2004)

[3] Interaction of hydrogen with nitrogen interstitials in wurtzite GaN, A.F.Wright, J.Appl.Phys. 90 **6526 - 6532** (2001)

[4] Classical simulations of the properties of group-III nitrides, J.A. Chisholm, D.W.Lewis, P.D.Bristowe, J.Phys.Condens.Matter 11 **L235 - L239** (1999)

[5] Diffusivity of native defects in GaN, S.Limpijumnong, C.G.van de Walle, Phys.Rev.B 69 **035207** (2004)

[6] Vacancies in wurtzite GaN and AlN, K. Laaksonen, M. G. Ganchenkova, R. M. Nieminen, J. Phys. Condens. Matter 21 **015803** (2009)

[7] First-principles calculations for defects and impurities: Applications to III-nitrides, C. G. Van de Walle J. Neugebauer, J.Appl.Phys. 95 **3851 - 3879** (2004)

[8] Vacancies in wurtzite GaN and AlN, K. Laaksonen, M. G. Ganchenkova, R. M. Nieminen, J. Phys. Condens. Matter 21 **015803** (2009)

[9] Nitrogen vacancies as major point defects in gallium nitride, M.G.Ganchenkova, R.M.Nieminen, PRL 96 **196402** (2006)

7 Solid solutions

7.1 Introduction

The alloys of AlN, GaN and InN have many interesting properties and have been widely used in engineering applications. In particular, growing the alloys under careful control of composition allows for engineering the band gap and associated electrical properties, which are relevant in III-V nitride devices (see [1], [2], [3], [4] and [5]).
In this section, the interatomic potentials presented earlier are employed to study the alloys over a range of compositions. Two approaches are taken. In the mean field formalism, each cation site is fractionally occupied by an "average cation" of two different species with the occupancies summing to unity. Secondly, we utilise the SOD (Site Occupancy Disoder, [8]) software to generate ordered structures with desired compositions explicitly and evaluate their free energies and other relevant quantities.
Previous work suggests that strain plays an important role in the stabilisation of the III-V alloys, suppressing phase separation. The effects of strain are also studied in this chapter.

7.2 Mean Field Approximation

7.2.1 Methodology

In the mean field approximation, the cation sites are occupied with varying fractions of Al, Ga and In to simulate the full range of ternary III-V nitride alloys. This formalism was used in previous work, such as [6], and is also discussed in [7].
The sites with fractional occupations in the GULP software are interpreted as being occupied by a species whose interatomic interactions are defined as a weighted average of the interactions of the distinct species that fractionally occupy the site. For example, in $Al_xGa_{1-x}N$, the cation-N interaction in the

mean field approximation can be written as

$$V_{\text{cation-N in alloy}}(r) = xV_{\text{Al - N}}(r) + (1-x)V_{\text{Ga - N}}(r) \tag{92}$$

where $V_{\text{Al - N}}$ and $V_{\text{Ga - N}}$ are the sums of Coulomb and short range interactions between the relevant cations and N in the binary compounds.
This method is a simplistic attempt to model a random alloy, where the probabilities of occupation of a cation site are determined by the composition of the alloy.
The starting point a and c parameters for the mean field optimisation calculations were the weighted averages of the structural parameters of the end members. The calculations were performed at zero temperature and pressure.
The mean field approach is most suitable for alloys in which the alternative cations are very similar in their ionic radii, as is the case for Al and Ga. In differs from the other two cations, resulting in much more strained alloys, less amenable to mean field treatment. $Al_xGa_{1-x}N$ alloys are considered here and compared with explicit calculations.

7.2.2 Results and discussion

The enthalpy and internal energy are equal in the following calculations as the temperature and pressure contributions are zero. The enthalpy of mixing of two binary compounds, ΔH, is calculated as

$$\Delta H(\mathrm{A}_x\mathrm{B}_{1-x}\mathrm{N}) = H(\mathrm{A}_x\mathrm{B}_{1-x}\mathrm{N}) - \big(xH(\mathrm{A}) + (1-x)H(\mathrm{B})\big) \tag{93}$$

where A, B are Al, Ga or In and H is the enthalpy.

The dependence of enthalpy and ΔH on alloy composition are shown in Fig. 49 and 50. The free energy is shown to vary approximately linearly with composition. At 0K the enthalpy of mixing is always positive, suggesting that the system is more stable when separated into the two constituent phases.

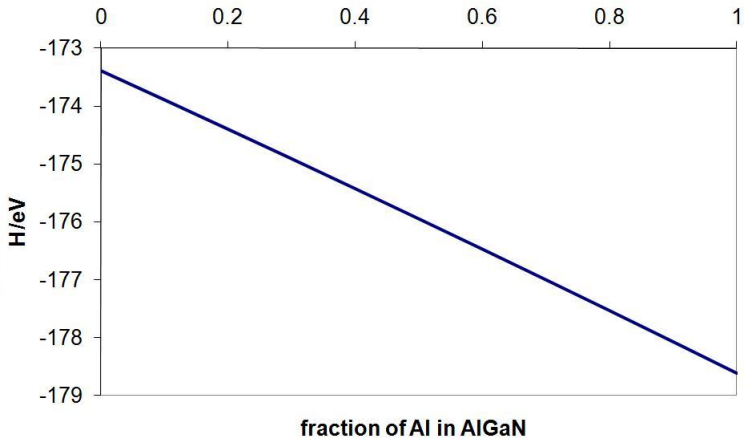


Figure 49: The enthalpy as a function of $Al_xGa_{1-x}N$ alloy composition, in the mean field approximation.

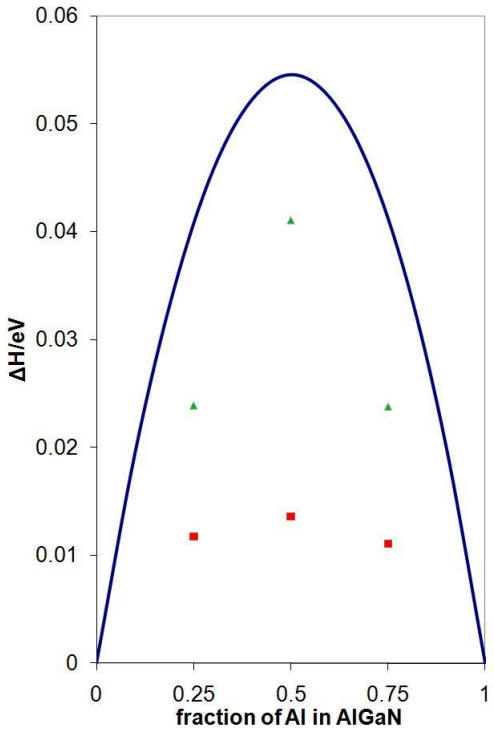


Figure 50: The enthalpy of mixing as a function of alloy composition, in the mean field approximation compared with lowest (red points) and highest (green points) explicit alloy results. For explicit structure calculations see section 7.3.

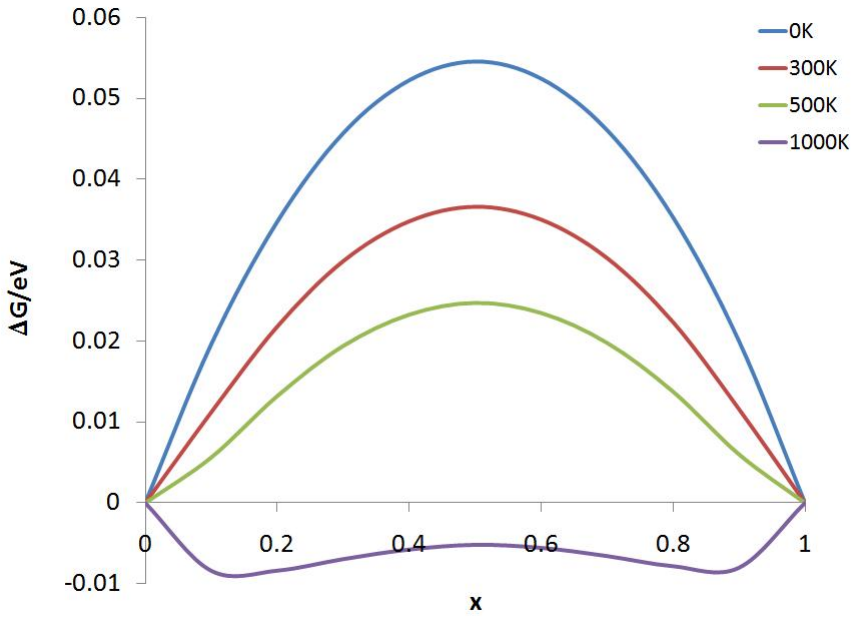
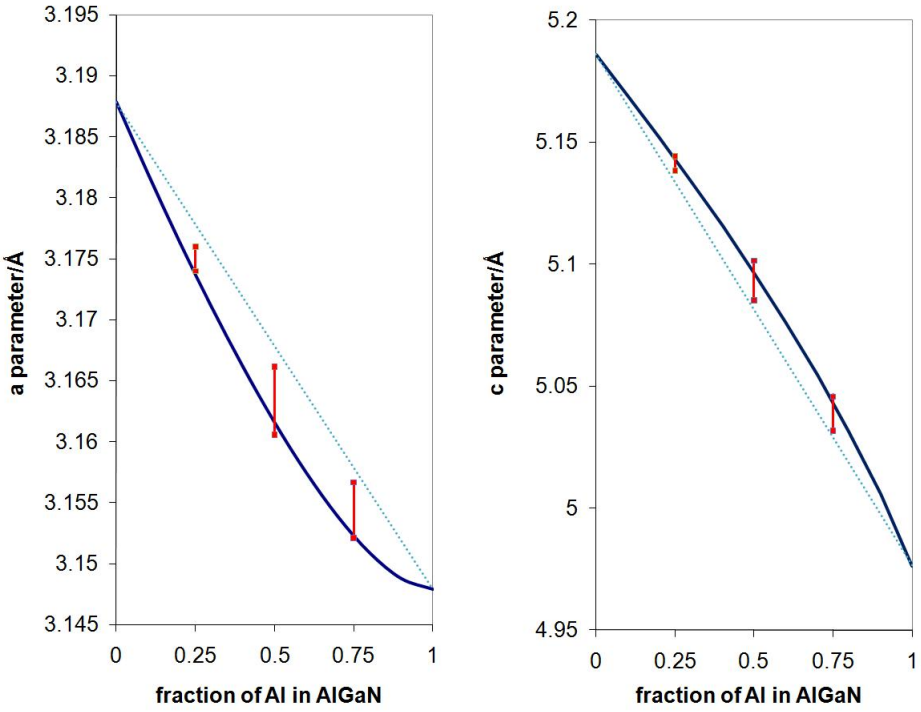


Figure 51: The free energy of mixing, assuming a fully disordered alloy for AlGaN at different temperatures.

The temperature dependence of the free energy of mixing, assuming a fully disordered alloy (which is the basis of the mean field approach; see Section 3.9.2) and taking into account only configurational, but not vibrational entropy, is shown in Fig. 51. The entropy contribution to the free energy was calculated via equation 80. The enthalpy of mixing becomes favourable at temperatures between 500 and 1000K.

The results are in qualitative agreement with early studies suggesting that the mixed alloys should be unstable, which would seem to be at odds with experiment where it has been shown that mixing does occur at finite temperatures. An important factor that is neglected in the present calculation is the epitaxial strain, which has previously been shown to reduce ΔH. A study of strained structures is presented in section 7.3. Another possibility is that experiment accesses a metastable configuration.

Next, the effect of composition on the cell parameters was studied. The results are shown in Fig. 52 and compared with Vegard's law, which postulates that the observable properties of an alloy vary as a weighted average of



(a) a parameter (b) c parameter

Figure 52: The structural parameter as a function of AlGaN alloy composition, in the mean field approximation (dark blue line). The light blue dashed line gives Vegard's law prediction. The red bars show the range of parameters obtained from explicit calculations from section 7.3.

the properties of the end members. For instance the structural parameter a can under this assumption be written as

$$a(\mathrm{A}_x\mathrm{B}_{1-x}\mathrm{N}) = xa(\mathrm{A}) + (1-x)a(\mathrm{B}) \tag{94}$$

The results of the explicit calculations from section 7.3 are also shown. The plots show significant deviations from Vegard's law. The a parameter is calculated to be smaller than the Vegard's law predictions, whereas c is larger. Both show a relatively smooth bowing out of the curves. The a parameter is at the lower end of the range from the explicit calculations and the c parameter at the higher end.

The mean field approach gives results which are reasonably close to those calculated with explicit calculations but it is unclear how reliably it models a random alloy. Overestimation of the enthalpy of mixing is a common problem since the method does not include possible energy lowering distortions [7].

7.3 Explicit cation ordering

7.3.1 Methodology

The mean field approximation is a useful starting point for the study of alloys but it is rather limited due to its simplistic nature. A more rigorous approach is to consider different explicit cation arrangements and compare their relative stabilities.

In the present work, the Site Occupancy Disorder program (SOD; see [8] and Section 3.9.2) was used to generate all inequivalent structures in $2 \times 2 \times 2$ and $1 \times 1 \times 8$ wurtzite supercells. Four cases were studied: $Al_xGa_{1-x}N$ with $x = 0.25, 0.5$ and 0.75 as well as $In_{0.25}Ga_{0.75}N$. In the case of the 25:75 mixtures the $2 \times 2 \times 2$ supercell has 38 inequivalent arrangements and $1 \times 1 \times 8$ has 116. In the 50:50 alloy the corresponding values are 190 and 810. The $1 \times 1 \times 8$ arrangements are by construction composed of layers of like cations (see Fig. 53). The labelling of the structures is arbitrary. The prefix Z indicates a $1 \times 1 \times 8$ supercell and prefix R indicates a $2 \times 2 \times 2$ structure in this Chapter.

The effect of temperature was studied via the inclusion of the degeneracy entropy S_{deg} of each arrangement, provided by SOD . Whilst at zero temperature the structure with the lowest internal energy is most stable, at higher temperatures this entropy may make a significant contribution to the reduced energy, E_{red}, as outlined in Section 3.9.2.

In addition to degeneracy entropy, there is a vibrational entropy contribution, which was not evaluated in this work due to the problems with modelling temperature dependences with the current potentials as explained in Chapter 5.

In modelling InGaN, a further issue arises. The spring constants employed

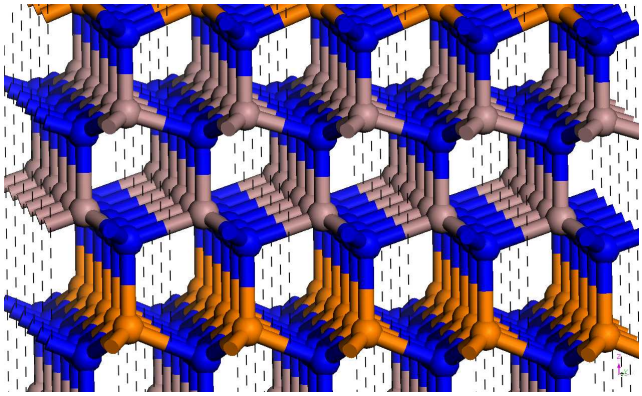


Figure 53: The lattice one of the configurations of $In_{0.25}Ga_{0.75}N$ with a 1×1×8 supercell. The dashed lines indicate the unit cell, which can be seen in Fig. 54. As can be seen from the diagram for any $1 \times 1 \times 8$ configuration the structure is by construction composed of layers of In (orange) and Ga (grey). The N ions are in blue.

in the models for GaN and InN are unequal in order to enable correct high frequency dielectric constants to be calculated. To simulate the alloy a spring constant which is a weighted average of the GaN and InN spring constants is used. Hence $k_2 = 42.828$ eV/Å^2 and $k_4 = 81250$ eV/Å^4. One could in principle improve on this procedure by changing the spring constant based on the number of like/unlike nearest cation neighbours and weight the spring constants accordingly, which is, however, beyond the scope of the SOD software at the moment. Another possible improvement would be to make the spring constant direction-dependent, so that the spring constant changes depending on the type of cation in a particular direction. This type of interaction is not implemented in the GULP code at present.

A more sophisticated treatment of polarisability would be to calculate the polarisability, and hence the spring constant, via the Madelung field (see Section 2.1). Ref. [9] suggested that polarisability might be environment dependent. [10] subsequently postulated that a simple relationship exists between the polarisability α_0 of an anion in a crystal with an anion Madelung field V_0 and an isostructural crystal with polarisability α' and Madelung field V'

$$\alpha' = \frac{\alpha_0 V_0}{V'} \tag{95}$$

If we know the Madelung field and polarisability of a nitride ion in one of the binary compounds that make up the alloy, and the Madelung field in the alloy, we can deduce the polarisability, and hence the spring constant (see Eq. 8), of the nitride ion in the alloy. The Madelung field of the alloy can be calculated with some starting value of the spring constants, e.g. a weighted mean of the end members, by optimising the alloy structure in GULP. A new value of polarisability, and a new spring constant, can then be calculated via Eq. 95. The alloy structure can be optimised with the new spring constant and a new Madelung field calculated until convergence is achieved in this iterative procedure. A scheme like this may be implemented in the future to improve on the present methodology.

The weighted average spring constants, used in this work, were tested in the pure binary systems (i.e. GaN and InN) to ensure that unphysical behaviour does not occur when they are used. The GaN structural parameters are well reproduced (a = 3.187Å, c = 5.188 Å). The high frequency dielectric constants are higher than those calculated with the original GaN spring constants (ϵ_{11} = 6.28 and ϵ_{33} = 6.37) and there are no imaginary phonon frequencies for the wurtzite structure, which indicates that a true minimum rather than a saddle point has been found.

For InN, the structural parameters with the altered spring constant are a = 3.570Å, c = 5.697 Å and the high frequency dielectric constants ϵ_{11} = 4.55 and ϵ_{33} = 4.64. No imaginary frequencies appear for wurtzite.

To simulate epitaxial strain in the material in the ab plane, the a parameter was fixed at 3.19 Å, approximately that of the optimised GaN structure, whilst other parameters, including the fractional coordinates within the unit cell, were allowed to relax. E_{red} was calculated by comparing with the energies of similarly strained binary alloys, i.e. with the a parameter fixed at 3.19Å. This model assumes that the a parameter of the substrate does not change in the presence of the epitaxial layer. It also assumes there are no strain-relieving dislocations.

Ternary III-V alloys are most frequently grown as epitaxial thin films. In the current model we assume an infinite extent of the lattice in the c-direction, simulating a thick layer. Thick layers are likely to have a higher concentra-

tion of dislocations that relieve epitaxial strain and lead to phase separation [12]. To improve the model, surface effects would have to be included as well. Furthermore, one could include a direct interface between the substrate and the epitaxial layer. This interaction is at present confined to fixing the a parameter of the alloy supercell, which is a rather crude approximation.

7.3.2 Results and discussion

$In_{0.25}Ga_{0.75}N$

ΔE_{red} for the different configurations explored (referenced by arbitrary numbers, $2 \times 2 \times 2$ supercells prefaced by R and $1 \times 1 \times 8$ by Z) is shown in Fig. 56 on p. 169. A summary of the lowest energy configurations at different temperatures (including the effect of degeneracy entropy), with and without epitaxial strain are shown in Table 12. Configuration R15 (see Fig. 54) is in all cases the one lowest in energy. The highest energy configurations R16 and Z112, are in fact equivalent (see Fig. 54 and 62). For $T = 0K$ structure Z150 (Fig. 54) is the highest in energy.

Structures Z150 and R16 are similar layered arrangements. In the R16 arrangement 3 layers of Ga are followed by 1 layer of In along the z-direction. Z150 has a layer sequence 1(In):2(Ga):1(In):2(Ga):1(In):2(Ga):1(In):6(Ga), a somewhat less symmetrical arrangement than the others.

Our results are in good agreement with DFT calculations in [14], which show that a structure equivalent to configuration R15 is the lowest in energy and structure R16 is the highest in energy. The configuration set in the previous work is less exhaustive than that presented here. The conclusion the authors draw is that cations of the same species are preferentially distributed as far from each other as possible. This finding can potentially be explained by the fact that a regularly distributed arrangement relieves internal strain in the material.

While our bulk calculations predict [0001] superlattices to be energetically unfavourable with respect to other types of ordering, there has been some experimental evidence of ordering in [11] in the form of (0001) superlattice

Table 12: The lowest energy configurations for the $In_{0.25}Ga_{0.75}N$ alloy, both relaxed and under strain (with the a parameter kept fixed at 3.19 Å). The highest energy configurations are in brackets. The configurations are shown in Fig. 54 and 62. Note the arrangements R16 and Z112 are equivalent as both the $1 \times 1 \times 8$ and the $2 \times 2 \times 2$ supercell can produce an arrangement with single layers of In alternating with three layers of Ga.

	0K	100K	300K	600K
unstrained	R15 (Z150)	R15 (R16=Z112)	R15 (R16=Z112)	R15 (R16=Z112)
strained	R15 (Z150)	R15 (R16=Z112)	R15 (R16=Z112)	R15 (R16=Z112)

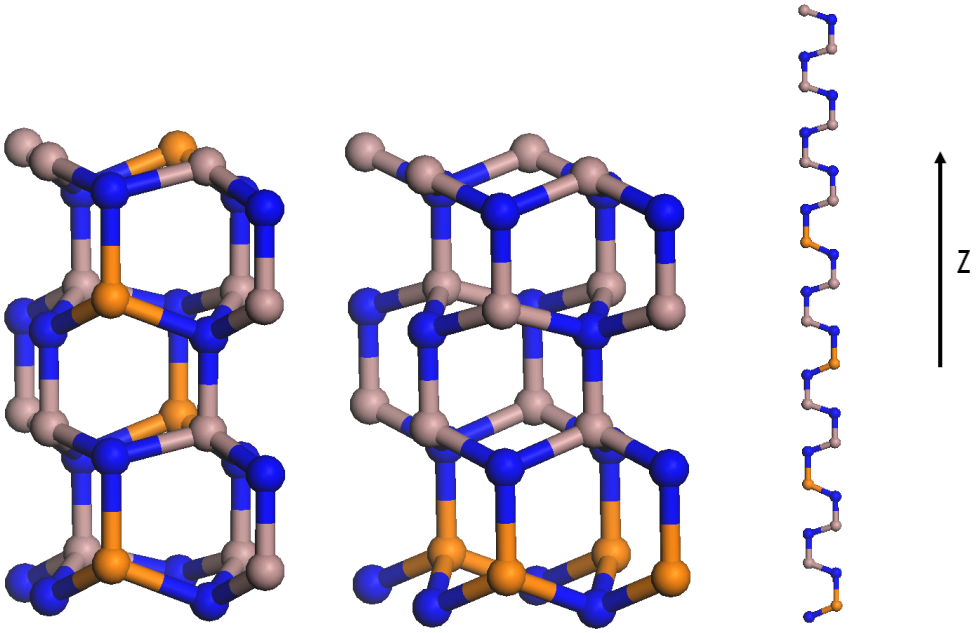
peaks, which are forbidden in a random alloy.

Reference [11] discusses experimental evidence of ordering in an $In_xGa_{1-x}N$. It suggests there is a competition between phase separation and ordering. Ordering is observed to dominate below In atomic fractions of 0.2 and phase separation is observed above those fractions. The ordering was suggested to occur along the c axis with separate planes of In and Ga ions.

It has been suggested that kinetic effects on the surface might lead to the stabilisation of the [0001] superlattices. A possible mechanism is discussed in [13]. This mechanism is used in [15] to explain how pyramidal ordering arises, i.e. ordering along the $[1\bar{1}01]$ surface in zincblende for which there is some experimental evidence in AlGaN [15]. This type of ordering, depicted in Fig. 55 was initially suggested for $In_{0.5}Ga_{0.5}P$ and is equivalent to ordering along the z-axis in wurtzite as explained in Fig. 10.

ΔE_{red} of the most stable configurations is shown in Table 13. The values are positive for all the temperatures under study in the unstrained case. Strain release has the dramatic effect of stabilising all the structures as ΔE_{red} becomes negative for most of the $2 \times 2 \times 2$ configurations.

The differences in ΔE_{red} of the lowest and highest configurations and the two lowest configurations is shown in Table 14. The differences between the lowest and highest energy configurations, of about 0.22eV per unit cell of wurtzite, are large relative to kT at 300K, which is approximately 0.03eV



(a) Config. R15 (b) Config. R16 (c) Config. Z150

Figure 54: Low and high energy configurations for $In_{0.25}Ga_{0.75}N$. The grey atoms are Ga and orange atoms are In. N is in blue. Configuration R16 and Z150 are very similar but the stacking of the layers is slightly different as discussed in the text.

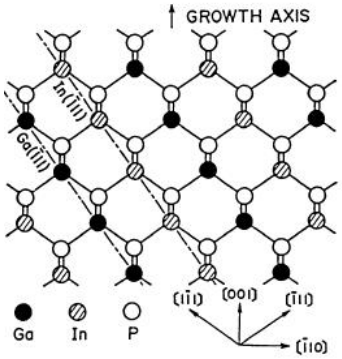


Figure 55: Pyramidal ordering as suggested for zincblende $Ga_{0.5}In_{0.5}P$ in [16]. This is a CuPt type arrangement.

Table 13: ΔE_{red} per unit cell of wurtzite for lowest energy configurations for $In_{0.25}Ga_{0.75}N$ in eV per wurtzite unit cell (i.e.containing two cations).

	0K	100K	300K	600K
unstrained	0.185	0.183	0.180	0.176
strained	-0.109	-0.110	-0.113	-0.118

Table 14: The differences in energy between the lowest and highest configurations for $In_{0.25}Ga_{0.75}N$ in eV per unit cell of wurtzite. The configuration reference numbers of the lowest and highest energy configurations are given in Table 12. The values in brackets are the differences between the two lowest energy configurations.

	0K	100K	300K	600K
unstrained	0.22 (0.061)	0.22 (0.058)	0.22 (0.053)	0.22 (0.045)
strained	0.20 (0.065)	0.20 (0.063)	0.20 (0.057)	0.20 (0.049)

per unit cell of wurtzite and means that the high energy configurations are very unlikely to be observed in the bulk at 300K. However the differences between the lowest two configurations are relatively small, about 0.06eV per unit cell of wurtzite. A number of configurations near the low energy end of the range are likely to be present.

The DFT results in [14] predict configuration R15 to have a formation energy of 56meV/cation. The present data are calculated per wurtzite unit cell, i.e. per two cations. The corresponding value from our work is around 70meV/cation. The highest energy configuration is predicted in [14] to have a formation energy of 138meV/cation as opposed to 200meV/cation with the present work. The differences between the DFT results and are calculations are of the order of the difference between the DFT and valence force field calculations (VFF) in [14]. However, our calculations give formation energies which are higher than the DFT results while the VFF results are lower.
The dependence of the reduced energy on the number of nearest like cation neighbours (i.e. In-In nearest neighbour pairs rather than In-Ga pairs) is

shown in Fig. 56. An interesting feature emerges. ΔE_{red} increases with increasing numbers of nearest neighbour In-In pairs for the $2 \times 2 \times 2$ supercells. Although there is a fair degree of scatter in the data the general trend is clear. The arrangement with the lowest ΔE_{red} is one where there are no nearest In neighbours for any of the In ions in the supercell. This pattern persists under conditions of pseudomorphic[11] epitaxial strain in the *ab* plane (see Fig. 57). In the $1 \times 1 \times 8$ supercells, each In cation has at least six other In ions as the nearest cation neighbours, which is not the case for $2 \times 2 \times 2$ supercells. The $1 \times 1 \times 8$ supercells have generally higher reduced energies, but the trend is opposite to the $2 \times 2 \times 2$ supercells and the reduced energy decreases as the number of like cation neighbours increases.

The highest ΔE_{red} for the $2 \times 2 \times 2$ arrangement coincides with one of the $1 \times 1 \times 8$ supercells. One can easily deduce that this will be the only $2 \times 2 \times 2$ supercell with In-only- and Ga-only- layers as shown by configuration R16 in Fig. 54.

These findings can be intepreted in terms of the increased internal strain induced in the alloy when In cations cluster. Due to the large lattice mismatch between Ga and In, the strain is minimised by ordering In in a staggered arrangement as shown in configuration R15 in Fig. 54. Small-scale clustering of In leads to strain at the interface of the In cluster and the Ga-rich portion of the alloy, rendering such arrangements energetically unfavourable.

This deduction is counterintuitive, since the positive sign of ΔE_{red} implies that phase separation, which is an extreme form of clustering of the two binary compounds, should occur. Once mixed, however, the alloys will assume an arrangement which will most effectively relieve the internal strain. The ordered structure with evenly dispersed In ions leads to strain release and the lowering of energy and is favoured over more clustered and strained structures with multiple interfaces between In- and Ga-rich regions.

[11] i.e. keeping the alloy structural parameters equal to those of the underlying substrate for the entire extent of the direction normal to the interface. In this case, the interface is in the *ab* plane and the substrate has parameters $a = b = 3.19$ Å and $\gamma = 120°$, which are kept constant for $-\infty < z < +\infty$ for the alloy epitaxial system. This is unlikely to occur in practice. Once the layer exceeds a certain critical thickness, the epitaxial strain would in practice be relieved by dislocations.

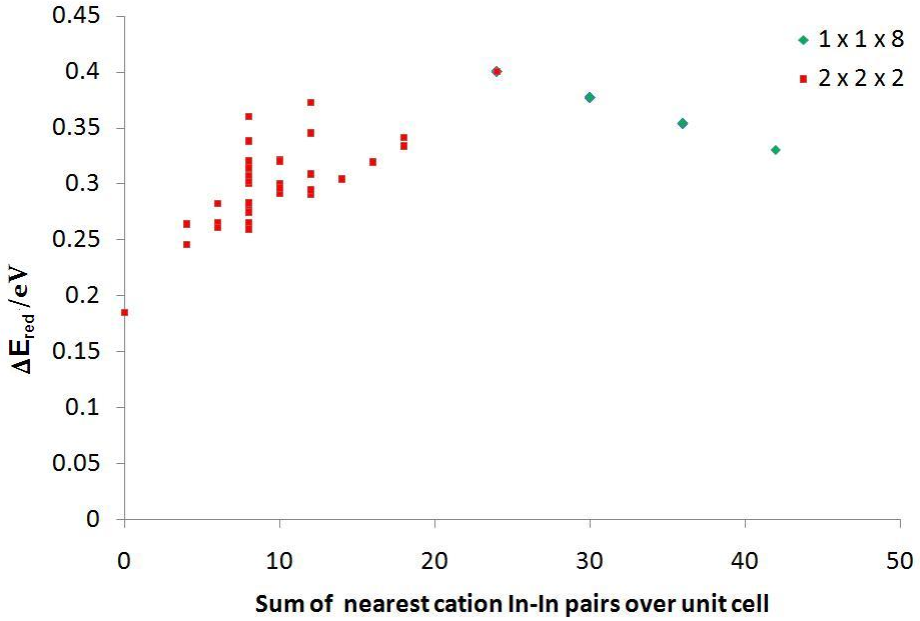


Figure 56: ΔE_{red} per unit cell of wurtzite vs. the sum of the number of nearest cation In-In pairs (summed over the unit cell).

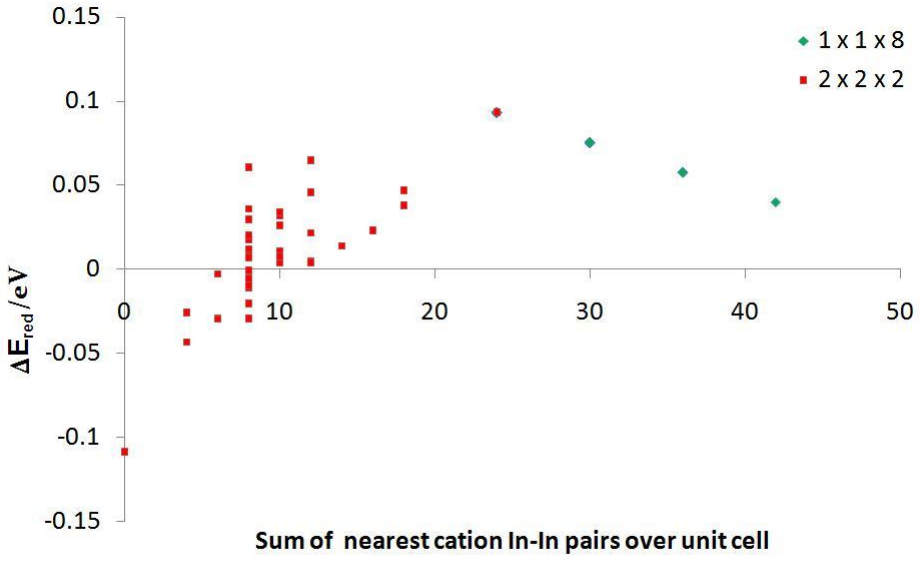


Figure 57: ΔE_{red} per unit cell of wurtzite vs. the sum of the number of nearest cation In-In pairs (summed over the unit cell) under strain.

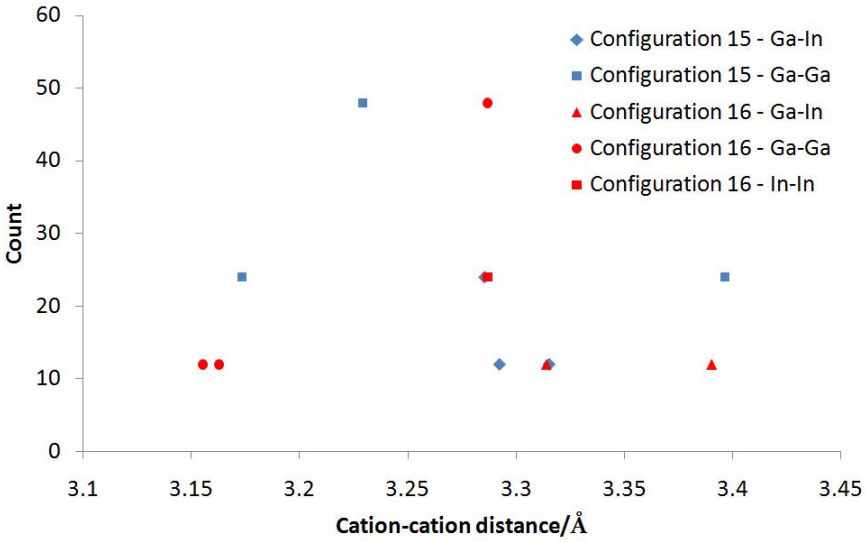


Figure 58: The distribution of cation-cation bond lengths in the low energy (configuration R15) and high energy (configuration R16) arrangements.

Detailed analysis of the GULP output suggests that the energy difference between configurations R15 and R16 arises 40% from the short range potentials and 60% from the Coulomb interaction. Since the charges in the Coulomb interaction are the same in the two cases, the difference has to be due to interionic distances. To understand the causes of the energy difference in the $2 \times 2 \times 2$ configuration, the weighted mean bond distances in the low energy (configuration R15) and high energy (configuration R16) were studied. Of course the weighted mean distance is only a crude way of estimating the effect of geometry on the lattice energy since the dependences are non-linear. Furthermore, second and further neighbours play a role. However, some qualitative insights can be gleaned from these considerations.

The cation-cation distances are shown in Fig. 58 and in Table 15. The cations interact with other cations via identical repulsive Coulomb interactions and a r^{-12} term, which has a very small effect. For the purposes of the energy calculation the cation-cation interactions of all combinations of cations may be treated as equivalent, i.e. as Coulombic interactions between charges 3+. It can be seen that configuration R15 does indeed have a slightly

Table 15: The weighted mean bond lengths in configurations R15 and R16 in $In_{0.25}Ga_{0.75}N$. Configuration R15 is the lowest energy configuration, R16 is the highest in energy. The results in the third column are from [17], calculated with the valence forcefield model with an $8\times8\times5$ supercell simulating a random alloy (i.e. cation sites are randomly occupied) and the bond lengths are averaged. The experimental results in the fourth column, from [18], are measured by total electron yield extended x-ray absorption fine structure. The samples were grown by molecular beam epitaxy and contain a mixture of wurtzite, zincblende and amorphous regions.

bond type	Config. R15	Config. R16	[17]	[18]
In-In	×	3.289	3.317	3.35
Ga-Ga	3.257	3.244	3.243	3.25
In-Ga	3.295	3.352	3.297	3.275
(cation-cation avg)	(3.270)	(3.274)		
In-N	2.140	2.083	2.126	2.09
Ga-N	1.965	1.978	1.963	1.94
N-N (shorter)	3.208	3.242	3.204	
N-N (longer)	3.494	3.512	3.471	
(N-N avg)	(3.280)	(3.280)		
N core-shell	0.0162	0.0256		

higher weighted mean distance between cations than the high energy configuration R16 suggesting decreased positive contributions from the cation-cation Coulomb interactions to $\Delta E_{\rm red}$.

The In-N bond lengths in configuration R15 are longer, suggesting that the short range repulsion will be reduced (but the Coulomb attraction is also reduced). The shorter Ga-N bond length in configuration R15 follows the opposite trend. The mean displacement of the shells from cores in R15 is shorter resulting in a less positive contribution. The N-N bonds are slightly shorter.

The results in Table 15 are compared with valence forcefield calculations simulating a random alloy, and experimental data with no clear ordering. The bond lengths are in good agreement with both sets of results although direct comparisons are difficult to draw as our simulations are based on ordered, rather than random, structures. Shorter bond lengths might be expected

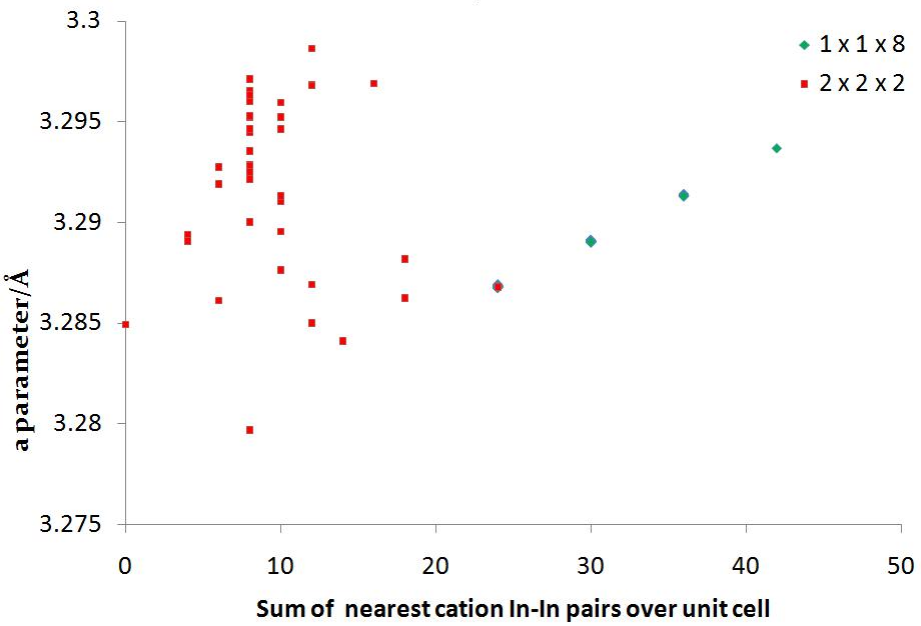


Figure 59: The a parameter as a function of number of nearest In-In neighbours in $In_{0.25}Ga_{0.75}N$.

due to the more efficient stacking of ordered structures, which is the case in some but not all bond lengths. The bond length differences between our calculations and other studies are within the range of error of our model (about 1%).

For the $1 \times 1 \times 8$ supercells the trend with the increasing numbers of like neighbours reverses. One can speculate that the clustering of layers means a reduction in the number of strained interfaces between layers of unlike cations and hence reduced overall interlayer strain.

Clustering of layers and the formation of [3,3] superlattices has been reported in [19] for zincblende $In_{0.5}Ga_{0.5}N$ by first principles and cluster expansion methods. Alternating groups of three InN and three GaN layers along the c-direction were predicted below the critical temperature of 1487K, above which disordered structures were predicted.

With larger supercells (extended in the c-direction) than in the present work, an even greater degree of layer clustering would be possible. For example, an arrangement with 24 Ga layers alternating with 8 In layers could be mod-

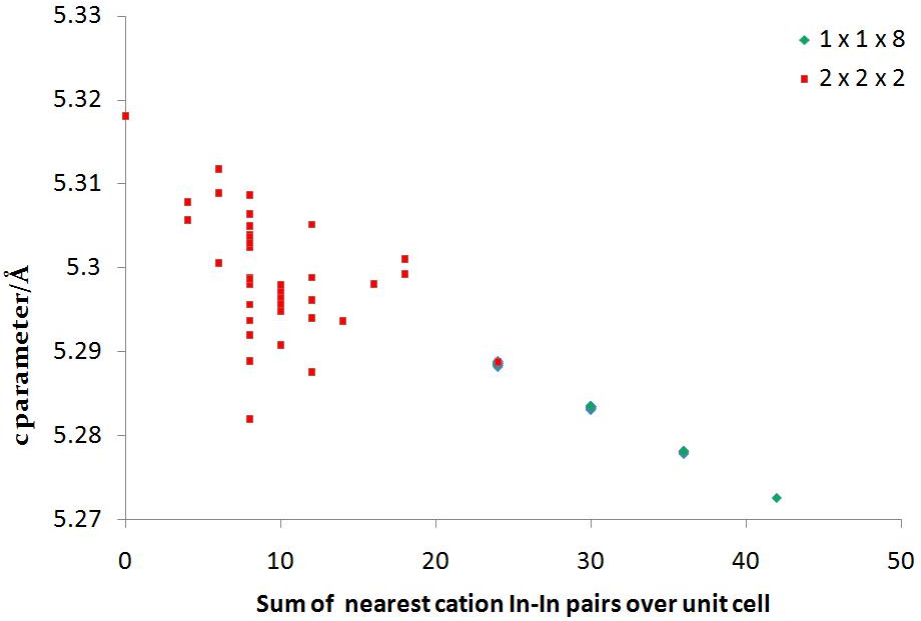


Figure 60: The c parameter as a function of number of nearest In-In neighbours in $In_{0.25}Ga_{0.75}N$.

elled. If the trends exhibited in our calculations are extrapolated, the system would be expected to undergo an even greater extent of clustering.

Finally, the structural parameters were plotted as a function of the number of nearest In-In neighbour pairs. The results are shown in Fig. 59 and 60. Again a clear relationship is observed: the a structural parameter increases with the number of nearest neighbour In-In pairs whilst the c parameter decreases.

The c parameter might decrease for the $1 \times 1 \times 8$ configurations due to the more efficient stacking of layers. If successive layers are of the same kind, their energies are minimised at the same interlayer distance. If they are of a different kind, a "compromise" distance is found. To relieve the strain of such an arrangement might require elongation in the c direction.

The a parameter is the lowest in arrangements where the In ions are interspersed between the Ga ions, which will result in the the smallest perturbation of the cation layers, which are predominantly Ga, and therefore the smallest increase in the a parameter relative to pure GaN. Future analysis

Table 16: The lowest energy configurations for the AlGaN alloy, both relaxed and under strain (with the a parameter kept fixed at 3.19 Å). The highest energy configurations are in brackets.

%Al	Unstrained			
	0K	100K	300K	600K
25	Z88 (R16=Z112)	Z88 (R16=Z112)	Z88 (R16=Z112)	Z88 (R16=Z112)
50	Z452 (Z191)	Z452 (Z191)	Z452 (Z191)	Z452 (Z191)
75	Z105 (R1)	Z105 (R4)	Z105 (R4)	Z105 (R4)
% Al	Strained			
	0K	100K	300K	600K
25	Z88 (R16=112)	Z88 (R16=Z112)	Z88 (R16=Z112)	Z88 (R16=Z112)
50	Z452 (Z191)	Z452 (Z191)	Z452 (Z191)	R175 (Z191)
75	R38 (Z147)	R38 (Z48)	Z105 (Z48)	Z105 (Z48)

could include the study of elastic constants of the alloys, the dependence of the unit cell volume on the number of nearest In-In neighbours and the effects of defects on the bulk modulus of the alloys. $\mathbf{Al_xGa_{1-x}N}$

Next, $Al_xGa_{1-x}N$ will be considered for $x = 0.25, 0.5$ and 0.75. A summary of the highest and lowest energy configurations is shown in Table 16.

For $x = 0.25$ the lowest energy configuration is a layered structure with all the Al layers clustered together. The highest energy configuration is a layered structure with each Al layers interspersed with 3 Ga layers, which is configuration number Z112, or equivalently R16, in Fig. 62.

These patterns are reminiscent of the nearest neighbour dependences studied for InGaN. Here we plot ΔE_{red} against the number of nearest neighbour Al-Al pairs in Fig. 61. The same types of pattern emerge. The layered structures (i.e. $1 \times 1 \times 8$ supercells) are now however shifted lower in energy resulting in one of these structures with Al layers clustered together, having the lowest energy overall. The likely cause of the increased stability of the layered arrangements is the much smaller lattice mismatch relative to the InGaN case.

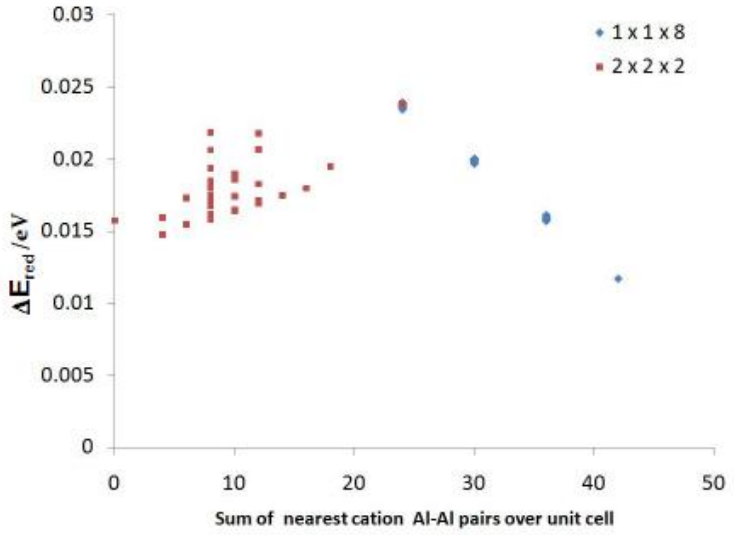
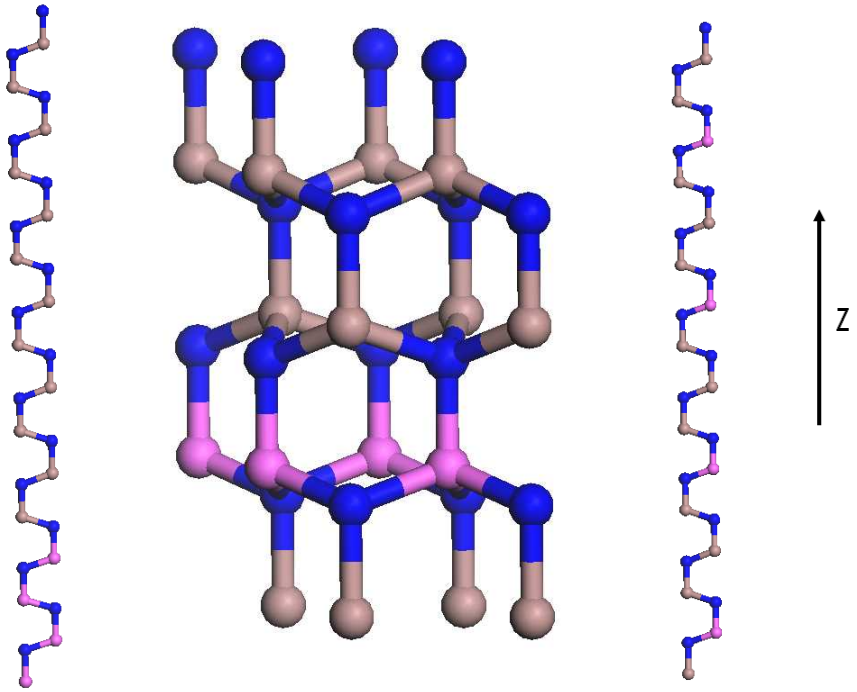


Figure 61: ΔE_{red} per unit cell of wurtzite vs. the sum of the number of nearest neighbour Al-Al pairs for $Al_{0.25}Ga_{0.75}N$.

Reference [20] studied $Al_xGa_{1-x}N$ where $x = 0.3 - 0.5$ and observed diffraction peaks forbidden in a random alloy. [21] attributes this pattern to alternating AlN and GaN layers along the c direction. 10:2 monolayer ordering was reported in AlGaN [22]. Certain growth conditions are thought to favour (0001) superlattices even though they are not necessarily energetically most favourable in the bulk (see [15] and references therein) which might explain why the structures predicted in the present work to have positive ΔE_{red} are stable in experiment. Another possibility is that the structures layered along the *c*-direction are metastable in the bulk. The discrepancy between experimental observations and the present calculations may also be due to the differing energetics at the growth surfaces relative to the bulk. Surface layers, esp. polar, can be expected to experience reconstructions, which might result in different structures being stabilised.

The calculation of diffraction patterns from the predicted structures is work in progress. Lowering of symmetry to the P3M1 spacegroup (no. 156) was observed for the Z112 (=R16) configuration.

The trends in the structural parameters in relation to the number of nearest Al-Al pairs differ somewhat from the InGaN case (see Fig. 63 and 64).



(a) Config. Z88 (b) Config. R16 (c) Config. Z112

Figure 62: The low and high energy configurations for $Al_{0.25}Ga_{0.75}N$. Pink atoms are Al, grey Ga and blue N. Configuration Z88 is the lowest in energy. Configurations R16 and Z112 are equivalent, and the highest in energy amongst all the configurations searched.

There is a less clear dependence between the number of like neighbours and a parameter although the $1 \times 1 \times 8$ arrangements have generally higher a parameters. The c parameter, as for InGaN, shows a decreasing trend with the number of like Al-Al pairs.

Values of ΔE_{red} per wurtzite unit cell for the lowest energy configurations are shown in Table 17. The values are much less positive than in the case of $In_{0.25}Ga_{0.75}N$. Just increasing the temperature to 600K, even in the absence of strain, is sufficient to turn ΔE_{red} negative. The effect of strain is much less pronounced than for InGaN, only yielding a reduction in ΔE_{red} of about 2meV per unit cell of wurtzite.

The differences between the highest and lowest, and the two lowest, calcu-

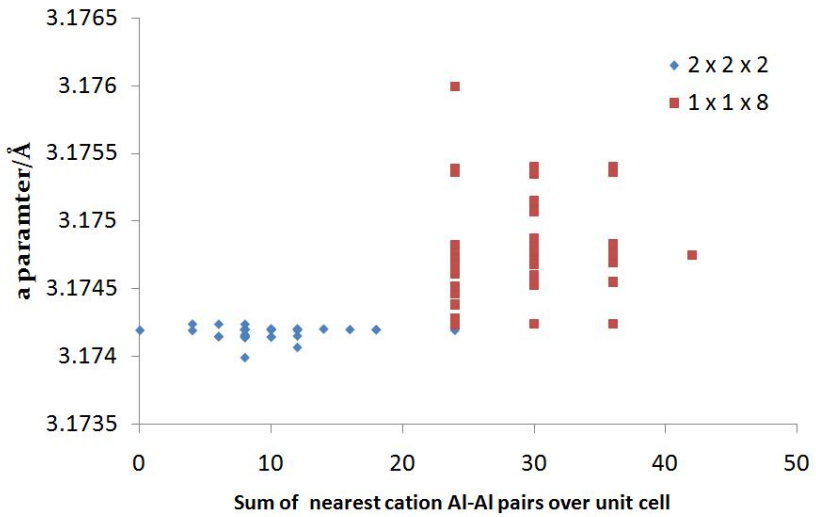


Figure 63: The a parameter as a function of number of nearest Al-Al neighbours in $Al_{0.25}Ga_{0.75}N$.

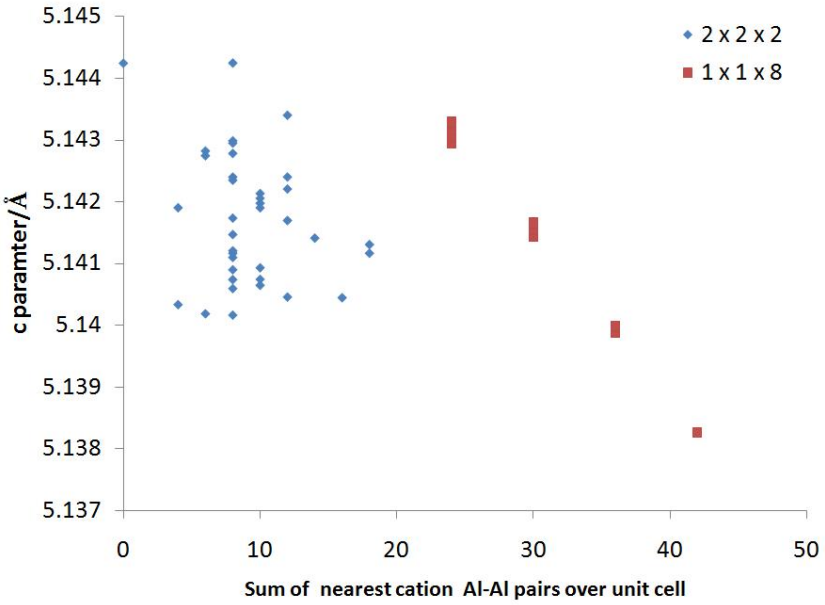


Figure 64: The c parameter as a function of number of nearest Al-Al neighbours in $Al_{0.25}Ga_{0.75}N$.

Table 17: ΔE_{red} per unit cell of wurtzite for the lowest energy configurations for AlGaN in meV per unit cell of wurtzite.

%Al	Unstrained			
	0K	100K	300K	600K
25	12	8	3	-6
50	14	11	5	-10
75	11	8	2	-7
% Al	Strained			
	0K	100K	300K	600K
25	10	7	2	-7
50	13	10	4	-9
75	5	4	3	-6

lated values of ΔE_{red} are shown in Table 18. The variations are significantly smaller than for InGaN. Even the difference between the highest and lowest configurations is of the order of 0.01eV per unit cell of wurtzite, which is comparable to kT at room temperature. It also means that the AlGaN case is more taxing for our model, since the small energy differences may well be within the range of error of the model.

The patterns in ΔE_{red} and the extent of its variation between the different configurations are similar for the other two Al fractions studied, 0.5 and 0.75, as can be seen from Tables 16 and 17. As for $x = 0.25$, the differences between configurations are small and the strain imposed does not stabilise the structures significantly.
The lowest and highest energy structures for $x = 0.75$ are shown in Fig. 65, 66 and 67. Configuration Z105 is the lowest energy configuration, apart from the strained structure at 0 and 100K. It is equivalent to the lowest energy configuration for $x = 0.25$, with Al layers clustered together. At low temperatures under strain, configuration R38, which is also the lowest energy arrangement for $In_{0.25}Ga_{0.75}N$, is the most stable. The high energy structures are mostly layered arrangements with Al layers alternating with Ga layers rather than clustered together.

Table 18: The differences in energy between the lowest and highest configurations for AlGaN in meV per unit cell of wurtzite. The values in brackets are the differences between the two lowest energy configurations.

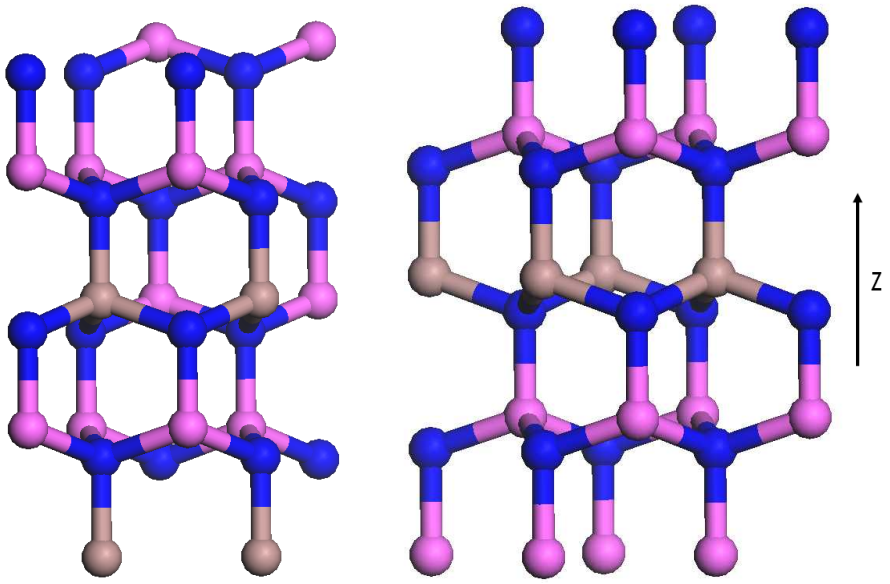
%Al	Unstrained alloy Energy difference meV per unit (cell of wurtzite) at temperature 0K	100K	300K	600K
25	12 (3)	15 (4)	21 (4)	29 (4)
50	27 (4)	30 (4)	34 (0.4)	46 (0.06)
75	13 (4)	15 (4)	21 (4)	29 (4)
%Al	Strained alloy Energy difference meV per unit (cell of wurtzite) at temperature 0K	100K	300K	600K
25	13 (3)	16 (4)	21 (4)	30 (4)
50	30 (5)	33 (5)	37 (2)	48 (0.02)
75	22 (0.4)	21 (0.2)	20 (0.6)	24 (5)

The results for $x = 0.5$ follow the same trends as for the 0.25 and 0.75 fractions in that the clustered layer arrangements are energetically favourable (see configuration Z452 in Fig.68 on p. 183). The alternating layer arrangement is the highest energy configuration. At high temperatures (600K) and under strain a more asymetrical arrangement 175 in Fig.68 is the most stable one. As was discussed earlier, the energy differences between the configurations are very small, of the order of 10^{-5}eV per unit cell of wurtzite at 600K.

Fig. 50 compares the explicit calculations with the mean field approach. ΔE_{red} calculated using the mean field approach is larger than the explicit calculations, as discussed in Section 7.2.2 on p. 161.

7.4 Summary and future work

To summarise, ternary III-V alloys were studied utilising the interatomic potentials derived in Section 5.1.1. Mean field and exhaustive enumeration in $1 \times 1 \times 8$ and $2 \times 2 \times 2$ supercells were employed. As in previous studies, the



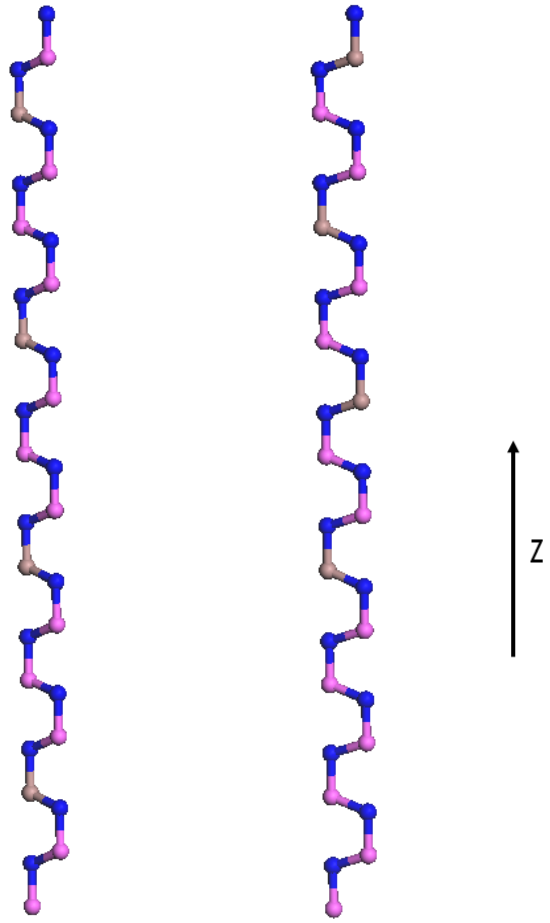
(a) Config. R1 (b) Config. R4

Figure 65: Two high energy configurations for $Al_{0.75}Ga_{0.25}N$. Configuration R4 is equivalent to Z48 shown in Fig 66.

current model predicts that unstrained alloys have a tendency to undergo phase separation. Epitaxial strain can have a strong stabilising effect, as was demonstrated for the case of $In_{0.25}Ga_{0.75}N$. Good agreement between a range of our and previous computational data is noted, for example in searching for the most stable configuration in the $In_{0.25}Ga_{0.75}N$ ordered structures set. We also find evidence of symmetry-lowering transitions.

In $In_{0.25}Ga_{0.75}N$, the lowest energy configuration has In ions maximally interspersed between Ga ions, presumably to relieve strain. The lowest energy configurations for AlGaN have clustered layers of Al and Ga. The optimised AlGaN structures have very similar energies, with differences of the order of kT at room temperature, so there is not a strong preference for the low-energy arrangements. In contrast, the energy differences in $In_{0.25}Ga_{0.75}N$ are larger.

The energetic and structural properties have been found to have a strong dependence on the number of nearest like cation-cation neighbours in the su-



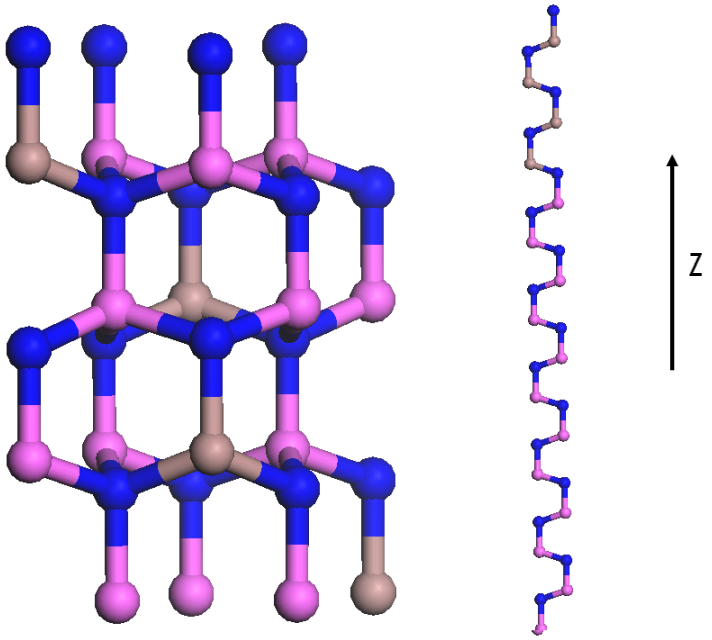
(a) Config. Z48 (b) Config. Z147

Figure 66: Two high energy configurations for $Al_{0.75}Ga_{0.25}N$.

percell, which can be related to the extent to which internal strain is relieved in the structure.

We now consider possible future directions of this work.

Thin films of the zincblende structured alloy have been reported to be stabilised by epitaxial strain when using cubic substrates such as 3C-SiC and GaAs [24]. There has been some suggestion, as outlined above, of pyramidal ordering along the $[1\bar{1}01]$ surfaces in zincblende. There is a large body of sim-

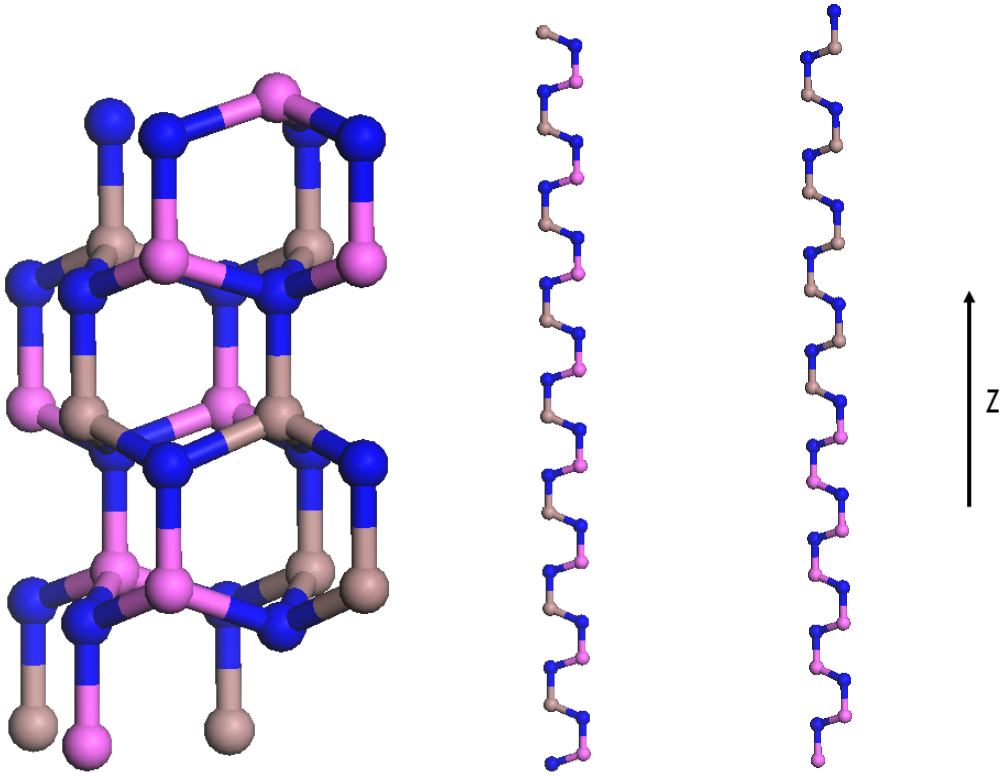


(a) Config. R38 (b) Config. Z105

Figure 67: The low energy configurations for $Al_{0.75}Ga_{0.25}N$.

ulations of zincblende ternary compounds (e.g. [23] and references therein). The present model would be well suited for the study of these systems. It would also be desirable to explore the configurations in larger supercells. At the moment one is limited by the SOD software in this as it breaks down if the number of configurations is too large.

Surface effects are of particular interest as the experimentally observed configurations might be determined by growth kinetics. The thermodynamics of the surface might favour different configurations from those that are energetically stable in the bulk, and those configurations might be "frozen in" unless the migration barriers for cations in the lattice are small enough. The energetics of migration therefore also warrants further study. This chapter has established an appropriate methodology for the study of these problems.



(a) Config. Z175 (b) Config. Z191 (c) Config. Z452

Figure 68: The low and high energy configurations for $Al_{0.5}Ga_{0.5}N$. This shows that the alloy prefers to phase separate unless it is under strain and a reasonably high temperature, in which case it assumes configuration R175.

References

[1] Dependence of the fundamental band gap of $Al_xGa_{1-x}N$ on alloy composition and pressure, W. Shan, J.W. Ager III, K.M. Yu, W. Walukiewicz, E.E. Haller, M.C. Martin, W.R. McKinney, J. Appl. Phys. 85 **8505 - 8507** (1999)

[2] Compositional dependence of the strain-free optical band gap in $In_xGa_{1x}N$ layers, S. Pereira, M.R. Correia, T. Monteiro, E. Pereira, E. Alves, A.D. Sequeira, N. Franco, Appl. Phys. Lett. 78, 2137-9 **2137 -**

2139 (2001)

[3] Universal bowing in group-III nitride alloys, J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager, S.X. Li, E.E. Haller, H. Lu, Solid State Commun. 127 **411 - 414** (2003)

[4] Superior radiation resistance of In1-xGaxN alloys: Full-solar-spectrum photovoltaic material system, J. Wu, W.Walukiewicz, K. M.Yu, W.Shan, J. W.Ager, E. E.Haller, Hai Lu, W.J.Schaff, W. K.Metzger, S.Kurtz, Journal of Applied Physics 10 **6477 - 6482** (2003)

[5] The composition dependence of the $In_xGa_{1x}N$ bandgap, K. P. ODonnell, I. Fernandez-Torrente1, P. R. Edwards, R. W. Martin, Journal of Crystal Growth 269 **100 - 105** (2004)

[6] Bulk and surface simulation studies of $La_{1-x}Ca_xMnO_3$ M.J.Akhtar, C.R.A.Catlow, B.Slater, A.M.Walker, S.M.Woodley, Chem.Mater.18 **1552 - 1560** (2006)

[7] `https://projects.ivec.org/gulp/help/gulp3.0_manual.pdf`

[8] Symmetry-adapted configurational modelling of fractional site occupancy in solids, R. Grau-Crespo, S. Hamad, C. R. A. Catlow, N H de Leeuw, J.Phys.Condens.Matt. 19 **256201** (2007)

[9] Calculations of Off-Centre Displacements of Divalent Substitutional Ions in CaO, SrO and BaO from Model Potentials, M.J.L. Sangster and A M Stoneham Phil. Mag. B43 **597-608** (1980)

[10] Potential models for ionic oxides, G.V.Lewis, C.R.A.Catlow, J.Phys.C:Solid State Phys. 18 **1149 - 1161** (1985)

[11] Article "Study of Phase Separation and Ordering in InGaN Alloys" by S.N.Basu in III-V nitride materials and processes III By T. D. Moustakas, S. E. Mohney, S. J. Pearton, Electrochemical Society. Dielectric Science and Technology Division, Electrochemical Society

[12] Phase separation in InGaN thick films and formation of InGaN/GaN double heterostructures in the entire alloy composition, R. Singh, D. Doppalapudi, T. D. Moustakas, L. T. Romano, Appl.Phys.Lett.70 **1089 - 1091** (1997)

[13] Surface energetics, pit formation, and chemical ordering in InGaN alloys, J.E.Northrup, L.T.Romano, J.Neugebauer, APL 74 **2319 - 2321** (1999)

[14] Limits and accuracy of valence forcefield models for $In_xGa_{1-x}N$ alloys, F.Grosse, J.Neugebauer, Phys.Rev.B 63 **085207** (2001)

[15] Pyramidal plane ordering of AlGaN alloys, M.Benemara, L.Kirste, M.Albrecht, K.W.Benz, H.P.Strunk, APL 82 **547** (2003)

[16] Large (6deg) off-angle effects on sublattice ordering and band gap energy of $Ga_{0.5}In_{0.5}P$ grown on (001) GaAs substrates, A.Gomyo, S.Iijima, I.Hino, Jap.J.Appl.Phys.28 **L1728 - L1730** (1989)

[17] Predicted bond length variation in wurtzite and zinc-blende InGaN and AlGaN, T.Mattila, A.Zunger, J.Appl.Phys. 85 **160 - 167** (1999)

[18] EXAFS study of groupIII nitrides N.J.Jeffs, A.V.Blant, T.S.Cheng, C.T.Foxon, C.Bailey, P.G.Harrison, A.J.Dent, J.F.W.Mosselmans, Mat. Res. Soc. Symp. Proc. 512 **519 - 524** (1998)

[19] Phase separation in group-III nitride alloys. L.K.Teles, M.Marques, L.M.R.Scolfaro, J.R.Leite, Braz.J.Phys. 34 **B593 - 597** (2004)

[20] Long range order in $Al_xGa_{1-x}N$ films grown by molecular beam epitaxy, D.Korakakis, K.F.Ludwig Jr., T.D.Moustakas, APL 71 **72 - 74** (1997)

[21] Simultaneous phase separation and basal-plane ordering in $In_xGa_{1-x}N$, M.Shimotomai, A.Yoshikawa, APL 73 **3256 - 3258** (1998)

[22] Evidence for multiple chemical ordering in AlGaN grown by metalorganic chemical vapor deposition, P. Ruterana, G. De Saint Jores, M. Laügt, F. Omnes, E. Bellet-Amalric, APL 78 **344 - 346** (2001)

[23] Ground state structure of coherent lattice-mismatched zincblende $A_{1-x}B_xC$ semiconductor alloys (x=0.25 and 0.75), S.Chen, X.G.Gong, S-H.Wei, Phys.Rev.B 77 **073305** (2008)

[24] Ab initio study of structural parameters and gap bowing in zincblende$Al_xGa_{1-x}N$ and $Al_xIn_{1-x}N$ alloys, M. B. Kanoun, S. Goumri-Said, A. E. Merad, H. Mariette, J.Appl.Phys.98 **063710** (2005)

8 QM/MM study of GaN

8.1 Introduction

In this chapter, the GaN potential developed in the previous chapters is combined with quantum mechanical approaches to calculate the electronic and structural properties of native point defects in the material. For details of the theory underpinning QM/MM methodology, see Section 2.3.

In our QM/MM approach, the system is modelled as a spherical cluster of radius 30 Å, with point charges simulating the remainder of the infinite crystal. The cluster is partitioned into the electronically active part, treated at the QM level (see Fig. 69), Region I. The defect is positioned approximately centrally in the QM region. The interface between the QM and MM parts, Region II contains Ga^{3+} ions represented by effective core potentials (ECPs; see section 2.2.2 on p. 54), designed to model the short range repulsion between the electrons originating from the QM region and the boundary ions. The ECPs prevent the electrons "spilling" over from the QM region into the MM fragment. The rest of the system is treated at the MM level of approximation, and is further divided into an active Region III, where cores and shells are allowed to relax with respect to their Cartesian coordinates, and an inactive Region IV, held fixed at the equilibrium geometry as calculated by the bulk MM model. Region V contains the point charges.

This approach was pioneered in [2] in a code named ICECAP. A more recent implementation of cluster embedding methodology, ChemShell [3], was used in [1] and [4] for ZnO and [5] for MgO as well as in the present work.

In this chapter, the detailed aspects of QM/MM methodology and software will first be reviewed. The complexities of the set-up of a perfect (defect-free) cluster are described and the convergence of the calculated energy for the bulk ionisation potential is investigated. Next, the native defect formation energies, optical transition levels, geometries and electronic structures are calculated. These properties give an indication of the relative abundance of the defects and their luminescence properties.

GaN is usually found to be n-type and it has in fact proven exceptionally dif-

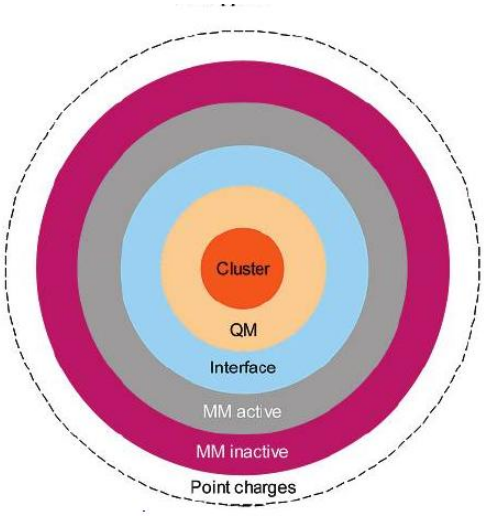


Figure 69: Regions in a ChemShell calculation. The cluster here denotes the defect region. The "cluster" and "QM" regions will be referred to as Region I. The interface in the present model contains Ga^{3+} ions whose valence electrons are described with a specially parameterised effective core potential. The interface is denoted Region II in the text. The MM active region will be referred to as Region III and the MM inactive region as Region IV. The point charges comprise Region V. From [1]

ficult to achieve p-type doping. Therefore, the behaviour at the conduction band minimum (CBM) is more relevant for interpreting experimental data. The knowledge of the defects near the valence band maximum (VBM) and their possible compensating properties might go some of the way towards explaining the difficulty with p-doping the material.

8.2 Methodology

8.2.1 Software and hardware

An implementation of the QM/MM approach, known as ChemShell [3] was used in this work. This software provides an interface between a variety of QM and MM codes, and implements a number of molecular mechanical methods (molecular dynamics, Monte Carlo), optimisers and analytical tools. The QM driver used in the present work is GAMESS-UK [6], [7]), and the MM driver was GULP, the same software that was used in the previous chapters

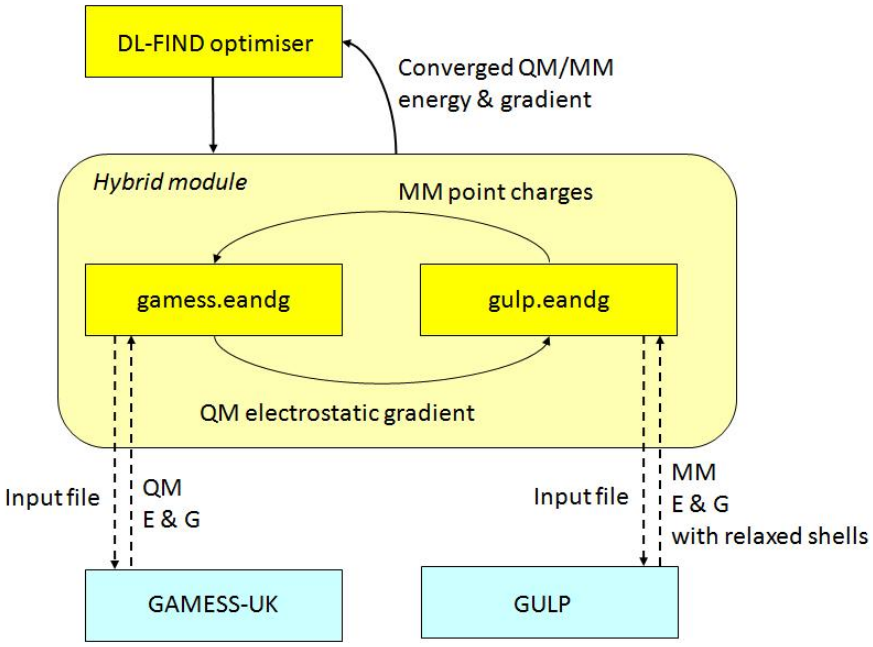


Figure 70: Chemshell schematic

in the development and validation of the interatomic potential.

To optimise the positions of the ions in the system, i.e. to determine the lowest energy configuration, a series of steps is performed at ionic geometries with successively lower energies, utilising an energy optimisation algorithm (see Section 2.4 on p. 60) analogous to a MM optimisation. At each configuration of ionic positions, the lowest energy of that configuration is determined by optimising the electronic distribution (using GAMESS-UK) and shell positions in the MM active region (using GULP), holding the ionic coordinates constant. The sequence of steps in a ChemShell calculation is shown diagrammatically in Fig. 70.

The calculation of the energies and forces on the QM region atoms is performed by GAMESS-UK at the starting geometry first. Once SCF convergence (convergence of the DFT calculation, see Section 2.2.2 on p. 50) is achieved, the output containing the QM energy and forces is termed gamess.eandg (=energy and gradient) in Fig. 70. This output also contains the electrostatic forces due to the QM region on the MM shells. With this information the shells in the MM-active region are relaxed, which results

in altered electrostatic forces on the QM region. Next, a new GAMESS-UK calculation is performed using the new shell coordinates. This sequence is repeated until the changes of the electronic distribution and shell displacements between successive calculations are below the desired threshold, i.e. until self-consistency is achieved between the QM and MM regions.
The converged forces on the ionic cores are then used by an optimiser, such as DL-FIND or newopt (see [3]), to find a new geometry with a lower energy. At this new set of ionic coordinates, another GAMESS-UK calculation followed by a self-consistent determination of shell positions is carried out. This process is repeated until desired convergence is achieved with respect to the total energy of the system, i.e. the energy minimum within a specified degree of accuracy is obtained.
The calculations in which the ionic configuration with the lowest energy is found will be referred to as optimisation runs. Calculations at a fixed ionic geometry with only electrons and shells being optimised are called single-point calculations, used in the determination of optical defect levels and ionisation potentials.

The QM/MM calculations are computationally expensive. They were performed on HeCToR, the UK National Computing Service phase 2a (Cray XT5h system) and 2b (Cray XT6). Full technical information is available in [8]. Optimisation calculations usually require many thousands of processor hours. Highly parallelised implementations of the ChemShell code were therefore used.

8.2.2 Calculating formation energies, ionisation potentials and defect levels

Some of the most important defect properties in semiconductors relate to the energetics of transitions between charge states of the system. The ionisation potential (IP) measures the energy required to remove an electron from the top of the highest occupied state, which can be the valence band or an in-gap defect level.
Both vacancies and interstitials can exist in a range of charge states. The

transitions between them relate to luminescence and their deep/shallow nature to their contribution to electrical conductivity (Section 3.8.2 on p.82). The defect formation energies affect their abundance (Section 3.8.1).
The IPs and optical defect levels calculated here are the values for a *vertical* transition, i.e. the energy of the ionised state is calculated at the geometry of the un-ionised system as a single-point calculation (see Section 8.2.1, p.190). In the case of the IP, we remove an electron from the VBM of a neutral perfect cluster, leaving it singly positively charged but do not allow the coordinates of the ions to relax from the equilibrium configuration of the neutral cluster. Such a calculation corresponds to a fast process where only the electrons and shells, which model electrons, are able to respond to the charge being removed. The ions, which are heavier, respond more slowly and can be approximated as stationary (i.e. in the Born-Oppenheimer approximation). In a system containing a defect, the electron may be removed from one of the in-gap states rather than the VBM.
In contrast to the *optical* transitions described above, the *thermodynamic* defect levels take the relaxation of the ions into account. These transitions assume that thermodynamic equilibrium has been reached and the most stable charge state is the one with the lowest formation energy. However, the activation energy for such a process may be prohibitively large, meaning that thermalisation might not occur on the timescale of an experiment. For a more detailed discussion of optical vs. thermodynamic defect levels, refer to Section 3.8.2, p. 82.
In calculating the IPs the electron is considered to be removed to infinity, i.e. the vacuum level (see Section 3.8.2), which is a convenient reference state, unaffected by the electronic changes associated with defects. It allows us to bring different calculations into the same energy frame.
As previously explained the QM/MM model only allows for polarisation of the QM, boundary and the MM active region. However, when a charged defect is introduced, the significant polarisation of the MM inactive region and the rest of the crystal is neglected. To correct for this inaccuracy, an additional term is introduced: this "Jost correction" [1] is derived by integrating the electrostatic energy over the surface of a sphere delineating the

boundary between Regions III and IV, R, which is set at 15Å in all cases in the present study, up to infinity. The Jost correction is then

$$E_{\text{Jost}} = \frac{q^2}{2R}\frac{\epsilon - 1}{\epsilon + 1} \tag{96}$$

where q is the charge of the defect and ϵ is the dielectric constant of the medium. The dielectric constant is taken from the bulk MM calculations in Chapter 5. For processes that occur fast so the atoms do not relax to new coordinates (vertical ionisation potentials and electron affinities), the high frequency dielectric constant is used in the calculation of the Jost correction. For thermodynamic processes (such as defect formation energies), the Jost correction is calculated using the static dielectric constant.
The IP can then be written as

$$E_{\text{IP}} = E_{\text{QM/MM}}^{\text{bulk charge 1+}}(X_0) - E_{\text{QM/MM}}^{\text{bulk charge 0}}(X_0) - E_{\text{Jost},\infty}(q = 1) \tag{97}$$

where $E_{\text{QM/MM}}^{\text{bulk charge 1+}}$ and $E_{\text{QM/MM}}^{\text{bulk charge 0}}$ are the QM/MM energies of a cluster as calculated by ChemShell and X_0 is the equilibrium geometry of the 0 charge state. $E_{\text{Jost},\infty}(q = 1)$ is the high frequency Jost correction for charge 1+.
In comparing the IPs with experimental values, one has to bear in mind that the electron is removed from the bulk. In experiment, surface effects and polarity are important. Non-polar surfaces have band structures similar to the bulk. Polar surfaces experience more bending of the bands. For these reasons direct comparison with experimental data has to be qualified.
The bulk electron affinities are more challenging to calculate with the current model because the strong delocalisation associated with electrons in the conduction band does not lend itself to cluster calculations.
Calculations of the quantity

$$E = E_{\text{QM/MM}}^{\text{bulk charge 1+}}(X_0) - E_{\text{QM/MM}}^{\text{bulk charge 1-}}(X_0) - 2E_{\text{QM/MM}}^{\text{bulk charge 0}}(X_0) \tag{98}$$

instead of the IP remove the problem with the definition the vacuum level. The above IP calculation refers to the process

$$\mathrm{GaN(bulk)} \rightarrow \mathrm{GaN(bulk)^{1+}} + \mathrm{e^-(vacuum)} \tag{99}$$

The optical defect transition levels are calculated similarly. For the 2+/3+ optical transition level of the N interstitial, for example, the following reaction equation applies

$$\mathrm{GaN(N_i^{2+})} \rightarrow \mathrm{GaN(N_i^{3+})} + \mathrm{e^-(vacuum)} \tag{100}$$

The energy corresponding to this transition, $E_{\mathrm{optical}}^{\text{vacuum level}}$, can be written as

$$\begin{aligned} E_{\mathrm{optical}}^{\text{vacuum level}} = &\left(E_{\mathrm{QM/MM}}^{\mathrm{N}_i^{3+}}(X_{\mathrm{N}_i^{2+}}) - E_{\mathrm{Jost},\infty}(q=3) \right) - \\ &- \left(E_{\mathrm{QM/MM}}^{\mathrm{N}_i^{2+}}(X_{\mathrm{N}_i^{2+}}) - E_{\mathrm{Jost},\infty}(q=2) \right) \end{aligned} \tag{101}$$

The first bracket refers to the energy of a system with an N interstitial in the 3+ charge state, at the optimised geometry of the 2+ charge state, $X_{\mathrm{N}_i^{2+}}$, including the Jost correction for charge 3+, $E_{\mathrm{Jost}}(q=3)$. The second bracket is the energy of a system with an N interstitial in the 2+ charge state, at its optimised geometry, including the Jost correction for charge 2+, $E_{\mathrm{Jost}}(q=2)$. This calculation is implicitly performed with respect to the vacuum level as the electron is removed to infinity. To determine $E_{\mathrm{optical}}^{\mathrm{VBM}}$, the energy with respect to the top of the valence band (see Section 3.7) we use the following equation

$$E_{\mathrm{optical}}^{\mathrm{VBM}} = E_{\mathrm{optical}}^{\text{vacuum level}} - E_{\mathrm{IP}} \tag{102}$$

Because both E_{IP} and $E_{\mathrm{optical}}^{\text{vacuum level}}$ are calculated with respect to the common vacuum level, this subtraction gives the energy of the optical transition level with respect to the VBM. E_{IP} is taken a positive number.
The energy with respect to the VBM then corresponds to the following reaction

$$\mathrm{GaN(N_i^{2+})} \rightarrow \mathrm{GaN(N_i^{3+})} + \mathrm{e^-(VBM)} \tag{103}$$

This calculation implies the electron is removed to the defect and placed at the top of the valence band, which is a reasonable assumption in a p-type material: p-doping in GaN is difficult to achieve and the material is commonly found to be n-type. The optical transition level with respect to the CBM is more relevant. This process corresponds to the equation

$$\mathrm{GaN}(\mathrm{N_i^{2+}}) \rightarrow \mathrm{GaN}(\mathrm{N_i^{3+}}) + \mathrm{e^-(CBM)} \tag{104}$$

and its energy may be calculated via

$$E_{\mathrm{optical}}^{\mathrm{CBM}} = E_{\mathrm{optical}}^{\mathrm{vacuum\ level}} + E_{\mathrm{A}} \tag{105}$$

where E_{A} is the electron affinity (as a negative number). Negative values of the defect level with respect to the VBM or CBM indicate that the level is located above the VBM or CBM, respectively. States *above* the CBM are resonances in the conduction band whereas states *below* the VBM are resonances in the valence band. As a consequence they have finite lifetimes and tend to decay rapidly.
One can similarly calculate the electron affinities of defects. For example, the affinity of the nitrogen interstitial in the charge state 2+ with respect to the CBM can be represented by the following reaction

$$\mathrm{GaN}(\mathrm{N_i^{2+}}) + \mathrm{e^-(CBM)} \rightarrow \mathrm{GaN}(\mathrm{N_i^{1+}}) \tag{106}$$

If we wish to calculate the optical transition level, the single-point calculation of the 1+ charge state is carried out at the optimised geometry of the 2+ charge state.
A negative value of a defect transition energy indicates that the implied process occurs spontaneously, which can be seen in Fig. 71 on p. 196. The black curve represents an initial state, i.e the left hand side in an equation such as Eq.104. Three final state (right hand side of a reaction) possibilities are shown. The first, orange curve, represents a lower-energy final state and energy E_1 is released in the optical transition by photoemission. Such a situation would arise if the defect state is a resonance in the conduction band,

in which case an electron would auto-ionise to the bottom of the conduction band, or if the state is a resonance in the valence band and captures an electron from it. The green curve represents a state which is thermodynamically stable with respect to the initial state, i.e. its minimum is lower than that of the initial state, but there is an activation barrier. For the optical transition from the initial to the final state to occur energy E_2 must be absorbed, e.g. in the form of photons. State 3, the blue curve, is thermodynamically unstable with respect to the initial state, i.e. the minimum of the curve lies above that of the initial configuration; E_3 is absorbed during the optical transition. In summary, the relative position of the energy minima of two charge states indicates their relative thermodynamic stability and hence their relative equilibrium concentrations. Positive vertical ionisation potentials and affinities indicate the amount of energy that needs to supplied in an optical process for a transition to occur. Negative optical defect levels indicate unstable resonances, i.e. electrons auto-ionising to the bottom of the conduction band or capturing electrons from the top of the valence band. The formation energies of the resonances are ill-defined.
After the excitation of the electron to a higher energy state, the energy may be dissipated as heat, i.e. phonons, or emitted in the form of light as the electron falls to a lower energy state, either its original state or another low-lying energy level. Photo-emission can provide useful experimental evidence of defect energetics.

The formation energies of defects can be calculated using Eq. 63. For instance, for a nitrogen vacancy in the charge state 3+ the formation energy refers to the following reaction

$$\mathrm{GaN(bulk)} \rightarrow \mathrm{GaN(with\ N_{va}^{3+})} + 1/2\mathrm{N_2(g)} + 3\mathrm{e^-(CB)} \tag{107}$$

where CB refers to electrons in the conduction band.
The energy of this reaction can be written as

$$\begin{aligned} E_F(\mathrm{N_{va}^{3+}}) = E_{\mathrm{QM/MM}}(\mathrm{with\ N_{va}^{3+}}) - E_{\mathrm{Jost},0}(q = 3+) + \\ +\mu(\mathrm{N}) - 3E_A - E_{\mathrm{QM/MM}}(\mathrm{perfect}) \end{aligned} \tag{108}$$

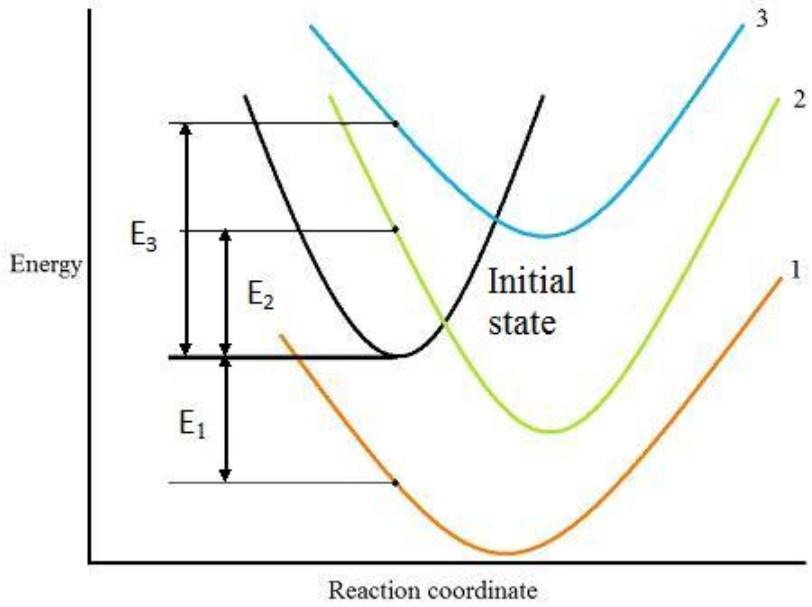


Figure 71: Optical transitions from the initial state (black curve). The energies with respect to the reaction coordinate are shown for different charge states, represented by different curves.

$E_{\mathrm{QM/MM}}(\text{with } \mathrm{N}_{\mathrm{va}}^{3+})$ and $E_{\mathrm{QM/MM}}(\text{perfect})$ are simply the results of the QM/MM calculation with and without the defect, with equal sizes of the QM region. The defect QM/MM calculation requires the appropriate static Jost correction for the 3+ charge state, $E_{\mathrm{Jost},0}(q = 3+)$. The electron affinity, E_A, accounts for each electron in Eq. 107 as energy is released when electrons move from the vacuum level to the lower energy level in the conduction band. The electron affinity E_A is derived from

$$E_{\mathrm{IP}} = E_{\mathrm{gap}} + E_A \tag{109}$$

where E_{gap} is the experimentally determined band gap at 4K[12], 3.505eV [9], and E_{IP} is the ionisation potential calculated with the same size of QM region as the defect and bulk calculations. A Jost correction for charge 1+ is applied to the ionisation potential calculation.

[12]There is a negligible error from not using a 0K value.

This semi-empirical approach, using experimental parameters to calculate defect energies, is necessary as our method does not reliably model the electron affinity. Placing an electron in the conduction band results in a delocalised state, which is poorly represented by cluster methods as the wavefunction is artificially limited by the size of the QM cluster.

In the nitrogen-rich limit, μ(N) can be easily obtained as one half of the energy of a N_2 molecule calculated with the basis set and functional used in the defect and bulk calculations. In the gallium-rich limit the condition in Eq. 65 in Section 3.8.1 must be satisfied. As was shown in Section 3.8.1, in the Ga-rich limit, $\mu(\mathrm{N}) = \mu(\mathrm{N}[\mathrm{N}_2]) + \Delta H_f[\mathrm{GaN}]$ where the latter term is the enthalpy of formation of GaN, experimentally determined as -1.17eV [11]. It is corrected for temperature to obtain the value at 0K in the same fashion as the cohesive energy of the Ga metal below, to -1.22eV.

In the gallium-rich conditions, the chemical potential of Ga is given by the cohesive energy of Ga metal, obtained from experimental measurements. The enthalpy of formation under standard conditions (i.e. at 298K) is 272.0kJ/mol (section 5.12 in [12]). The temperature dependence is taken from that of germanium, which next to Ga in period IV (section 5.2 in [12]) as Ga data was not available. For gaseous Ge, the difference $H^0_{298K} - H^0_{0K}$ is 7.4kJ/mol. The Ga chemical potental in gallium-rich conditions at 0K is therefore is $\mu(\mathrm{Ga}) = -2.8\mathrm{eV}$ at 0K. In nitrogen-rich conditions, the chemical potential would be determined by a procedure analogous to the N chemical potential in Ga-rich conditions explained above, i.e. $\mu(\mathrm{Ga}) = \mu[\mathrm{Ga(atom)}] + \Delta H_f[\mathrm{GaN}]$.

8.2.3 Set-up of a perfect cluster

The system is simulated with a spherical cluster of radius 30 Å. Beyond this radius the description is approximated by a small number of point charges calculated with the Construct software [1], [4].

There are several steps in the cluster generation process. Firstly, the bulk structure is relaxed to its equilibrium geometry using GULP and the parameters developed in the previous chapters. The structural parameters, including

shell positions from the optimisation calculation, are then used by the Construct utility to create a spherical cluster, radius 30Å. The Madelung field (see Section 2.1) of the remainder of the crystal (i.e. beyond 30Å in our case) is simulated with a number of point charges. Some cluster ionic charges near the surface of the cluster may also be adapted by Construct better to model the electrostatic forces. In this step, the atoms that will form the central QM region are also labelled. The QM regions can be visualised and adapted as deemed necessary for a particular calculation; for example one might wish to terminate the cluster with a particular type of ion. which will have to be performed manually. Furthermore the QM region has to be sufficiently large for the problem at hand and any defects should be positioned as centrally as possible in the QM region. In the final step Regions II,III and IV (see below) of the calculation are defined based on their radii. The boundary of Regions I/II is defined by the QM cluster, while the user-specified thickness of Region II determines the boundary of Regions II and III. The III/IV boundary is determined as a sphere with radius 15 Å relative to the origin of the cluster and corresponds to the boundary of the inner region in the Mott-Littleton calculations in Chapter 6, where slightly larger radii of between 18 and 21 Å were typically used. Fig.46 shows that at a radius of region 1 in the MM defect calculation of 15Å, convergence to about 0.1eV is obtained.

In the QM/MM calculation with the cluster thus generated, the central portion of the cluster which is treated at the QM level, Region I, is terminated with nitrogen ions, which are capped with Ga ions, Region II. The gallium ions in Region II interact with Region I via Coulomb interaction and a specially parameterised large core ECP (see Section 2.2.2) on the Ga ions in region II, designed to prevent electrons spilling from the QM cluster into the MM portion of the calculation. Region II orbitals are effectively frozen, which prevents unphysical charge distortions due to the mismatch between the QM and MM forces. Hence Region II acts as a buffer between the QM and MM regions and their alternative descriptions of the system to ensure continuity.

The rationale for using Ga ions in the buffer region is their lack of valence electrons relative to nitrogen ions, which means that it is easy to model

them as point charges with ECPs. Parameterising an ECP for a nitrogen ion would be a much more challenging task as nitrogen ions are much more electronically active. Since Ga ions are used in Region II, Region I has to be terminated with N ions.

Apart from the interaction of Region I ions with themselves and the ECPs of the Ga ions in Region II, all the interactions are based on the MM model described in Chapter 5.

The convergence of the energies with respect to the size of Regions I and II is discussed in more detail below. Five sizes of Region I were explored: 5, 19, 42, 74, and 116 QM atoms. The difference in Region I and Region II radii is around 3.5 Å. Manual changes were made to this cutoff as the clusters were cut so that each N in Region I only has Region I or Region II Ga as nearest neighbours. This restriction ensures that the valence electrons of nitrogen ions do not spill unphysically from the cluster due to insufficient short range repulsion from Region III.

Region III comprises the MM active part of the calculation while Region IV is treated at the MM level of approximation but is not relaxed with respect to Cartesian coordinates during a QM/MM calculation. Region IV ion cores and shells are fixed at the optimised coordinates given from a bulk MM calculation. Any polarisation of this region by a charged defect is included via the Jost correction (see Section 8.2.2 on p. 192).

The parameters of the MM interactions are practically identical to the values used in the previous chapters to calculate bulk properties, defect properties and solid solutions. The potential in Results chapters I-III are slightly refined by improving the constant shifts of the different layers of the potentials and the associated polynomial regions, better to model the phase transition pressures. The bulk-distance potential function parameters, e.g. the parameters of the Ga-N Buckingham potential are left unchanged. This version of the potential is tabulated in 46 in the Appendix.

The *hybrid functional* used was b97-2 [13], which generally outperforms earlier functionals in several aspects such as structural parameters, binding energies and ionisation potentials as well as polarisabilities and reaction barriers. The *basis sets* (see Section 2.2.2) and ECPs are listed in the Appendix. For

Ga in Region I the SBKJC basis set and ECP [14] were used, which resulted in a considerable decrease in computational cost relative to the Def2-TZVP basis set [15], which was used for N.
Some of the most diffuse functions were removed to reduce the computational cost. In the Ga SBKJC basis the most diffuse L basis function (with $\zeta = 0.0746100$) was removed. The diffuse functions make it difficult to contain the electron density within the QM region and increase the computational expense of the calculation. Another function was uncontracted to give two functions with $\zeta = 2.1230000$ and $\zeta = 0.1939000$. Uncontracting a basis function recovers some degrees of freedom lost by excluding another basis function. The SBKJC *small core ECP* was left unchanged.
The Region II ECP for Ga is based on a large core pseudopotential simulating 28 core electrons, Stuttgart RLC. This pseudopotential was re-parameterised for the purposes of the current model. The criterion used in the parameterisation was the spread between values of the energies of the 1s N orbitals in a bulk QM/MM calculation. In the absence of boundary effects, no energy spread should be observed, as in the bulk all nitrogen ions are symmetry-equivalent. The presence of the QM/MM boundary causes an energy spread of about 0.8eV in an optimisation calculation with 74 QM atoms. Such energy spread is the lowest obtained with a manual and computationally expensive optimisation process. An improved and less manual method is desirable if better results are to be obtained with less computational expense.
The nitrogen Def2-TZVP basis set was altered by removing the f function as this unnecessarily increases computational effort. Nitrogen f orbitals would only contribute to octupolar and higher polarisation terms, which are negligible for nitride ions. The d functions were retained as they are needed to model the polarisation response.
For the calculations of vacancies, a basis set centred on the defect centre (i.e. the original position of the removed atom) was included. This extra ghost basis set enters into the QM calculation in the same way as basis sets centred on actual atoms and provides an additional degree of freedom, which has been found to improve results in previous work on ZnO [16]. Such ghost basis sets are commonly used in linear combination of atomic orbitals

(LCAO) methods in the calculation of anionic vacancies and in the counterpoise correction method (e.g. [20]).

One of the main challenges of developing a successful QM/MM model is the need to match the QM and MM forces at the boundary between the regions. An incorrect parameterisation of one pair of interactions in a MM model can be compensated for by another pair hence giving a seemingly correct model. This will be revealed in a QM/MM calculation as geometric distortions at the boundary. Finding a suitable set of parameters thus requires identifying the particularly distorted QM atoms at the boundary and determining improved parameters of the interaction between them and the MM neighbours by carefully considering the forces on the QM and MM atoms. For example, a significant distortion of a boundary atom towards the centre of the QM cluster might indicate that the MM forces are insufficiently attractive or they are too repulsive compared to the QM forces, which might be due to the interaction with either cations or anions in the MM region.

Large net forces indicate a problem with the interatomic potential although the magnitude of the force that is satisfactorily small to give acceptably small boundary distortions is difficult to determine *a priori*. Lighter atoms such as N will typically respond with larger distortions to the same force, as could be expected from Newtonian physics. A determination of the initial forces combined with a geometry optimisation for a number of MM parameterisations (forces) provides guidance on the acceptable magnitude of the initial force.

Once the culprit interaction causing the distortions has been identified, the potential model with the problematic interaction adapted for use in QM/MM has to be checked in a pure MM calculation and the physical properties recalculated. The potential will likely need to be re-parameterised in GULP by fitting to experimental data again and the new potential checked in a QM/MM calculation for boundary distortions. In principle, one can proceed iteratively until convergence is achieved between fitting to experiment and matching QM and MM forces although in practice this is a very time-consuming task. In this work, only one cycle was performed based on an original MM potential, not reported here, which was adapted by the above-mentioned procedure to

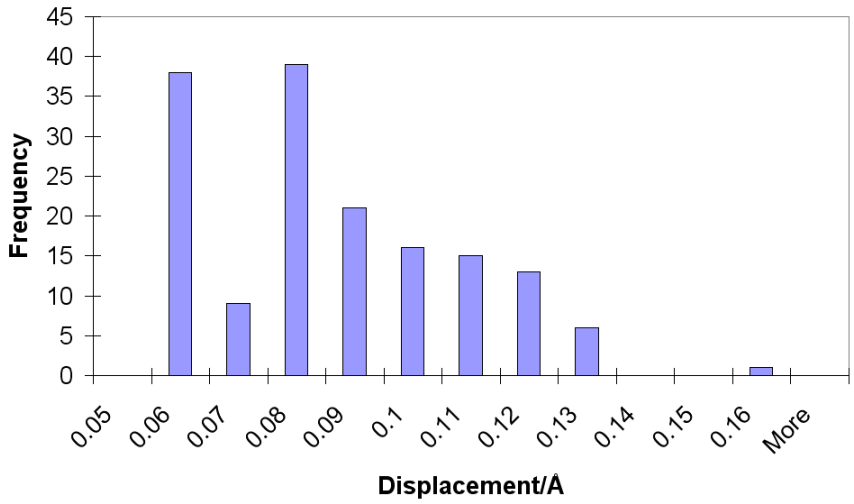


Figure 72: The atomic displacements for the 74-QM-atom cluster relative to their equilibrium positions in the MM model.

yield the parameters presented in the preceding chapters. It would be clearly desirable to make the process of finding suitable MM parameters less manual. While including the information from the QM forces improved the MM model in some respects (e. g. the signs of the piezoelectric constants), the errors on other observables increased (e. g. elastic constants).

The distribution of the displacement of ions from their position in the MM optimised bulk using a 74-QM-atom cluster is shown in Fig. 72. The y-axis indicates the number of atoms with their displacements from the MM equilibrium. All the displacements are less than 0.2 Å. One may argue that these are still relatively large compared to the structural parameters of GaN. However, providing that the QM cluster is large enough, the boundary distortions do not significantly interact with the defect centres in the middle of the cluster. The boundary effects are then present as perturbations of the Hamiltonian. Since all our energy calculations involve subtracting two QM/MM energies with the same QM region size, and hence the same starting boundary distortions, the perturbations should largely cancel out.

In addition to ensuring that the geometry is not significantly distorted, the electron density of the innermost QM ions should not exhibit marked dif-

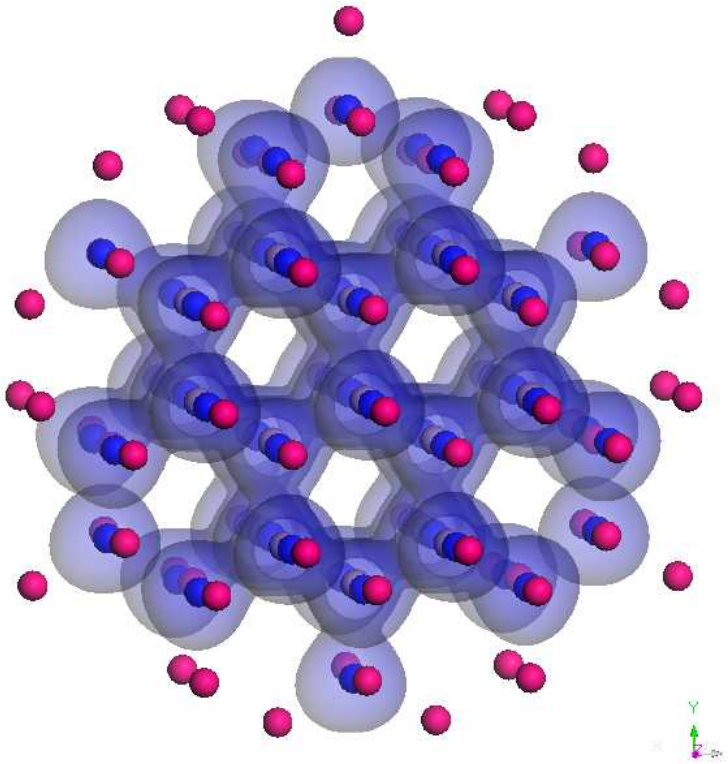


Figure 73: The charge density of the 74-QM-atom cluster. The blue atoms are N and grey are Ga in Region I. Pink atoms are Ga in region 2. The transparent blue clouds indicate the 0.05 e/Å^3 isosurface.

ferences from the boundary atoms. The charge density for the 74-QM-atom cluster is shown in Fig. 73. We did not attempt a quantitative assessment of the charge density variations between the centre and the edge of the cluster. However, a qualitative inspection does not reveal any gross charge distortions.

8.3 Results

In this section, calculations of ionisation potentials and the energetics, geometry and electronic structure of native defects will be presented.

8.3.1 Ionisation potentials

The calculation of ionisation potentials is relevant in two respects. Firstly, it allows for direct comparison with experimentally determined values. Secondly, this value is needed in the calculation of energetics of native defects later in this chapter via equations such as 108 and 109 if one wishes to keep the number of experimentally determined values used in the calculation to one, the band gap energy at 0K from [9].

The IPs calculated with the present model as a function of inverse QM region size are shown in Fig. 74. A trend of decreasing value of the ionisation potential with increasing QM region size is observed. The calculations were perfomed up to a QM-region size of 116 atoms, which with the present computational resources is the upper limit. The 5-QM-atom-cluster result (8.3eV) was excluded as it was deemed too small for a satisfactory calculation of the IP. Two least-squares lines were fitted to the plot of the IP against the inverse of the QM region size, one including all four calculated QM region sizes (19, 42, 74 and 116 atom) and one including only the largest three of these. As can be seen in Fig. 74, the two lines have slightly different y-axis intercepts, corresponding to the value of the IP in the limit of infinite cluster size. The intercept is 6.64eV when all four data points are included and 6.48eV with the three point subset. With this data it is difficult to draw firm conclusions about the precise value of the IP, especially since the number of data points is small. It should also be pointed out that energy convergence is usually plotted as a function of QM cluster radius (or the size of the supercell in PBC calculations). The number of QM atoms here is not a very accurate proxy for the radius of the cluster, as our clusters tend to be elongated rather than spherical to satisfy the requirements for N-termination of region II, for example. It is, however, clear that the IP is not fully converged even at our largest cluster size of 116 QM atoms.

A possible cause of inaccuracy in an IP calculation would be a highest occupied molecular orbital (HOMO; refer to Section 3.7) centred on an ion near the boundary of the QM region. Such a situation might arise, for example, if the system had a high degree of covalency. The broken bonds in the vicinity

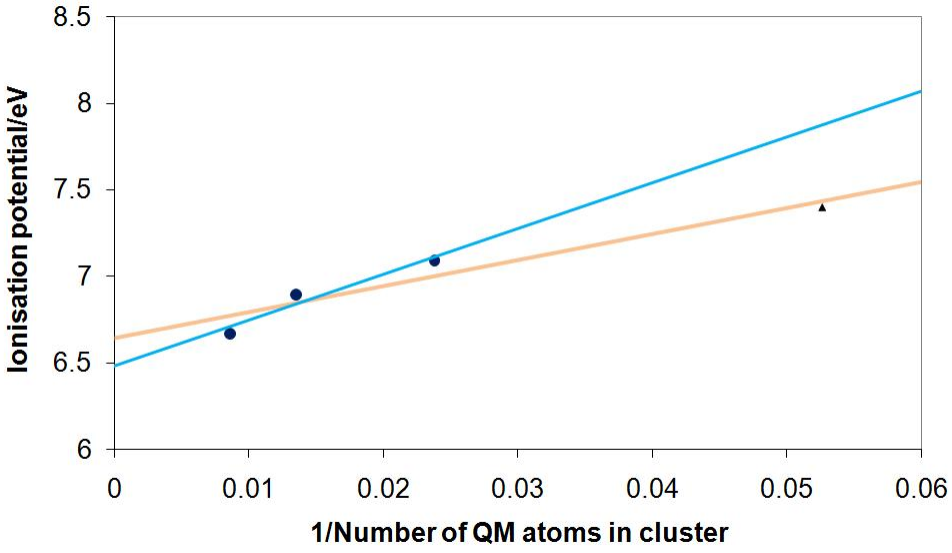


Figure 74: The ionisation potential of the perfect cluster as a function of the inverse of the QM cluster size. The black circles correspond to the results for the 116, 74 and 42 QM-region clusters and the black rectangle is the 19-QM-atom cluster result. Least-squares regression lines were fitted to the three circle points (blue line) and all four data points (pink line).

of the boundary would create states with high energy and might become the HOMO. However, one would expect such localised states to have energies which quickly converge with QM region size as their boundary environment is relatively independent of cluster size. Nonetheless, to exclude this possibility, the HOMO charge density was plotted and is shown in Fig. 75. The HOMO is quite delocalised with most of the density concentrated in p-like orbitals on nitrogen atoms intermediate between the centre of the cluster and the boundary, although there is a smaller but significant charge density on the boundary ions. Little of the charge density resides on the Ga ions or central nitrogens.

The boundary distortions tend to become slightly larger with increasing QM cluster size, possibly reflecting the mismatch between the equilibrium structural parameters of the QM and MM models, which might also contribute to the slow convergence.

At present, the issue of convergence of the IPs remains unresolved. In previ-

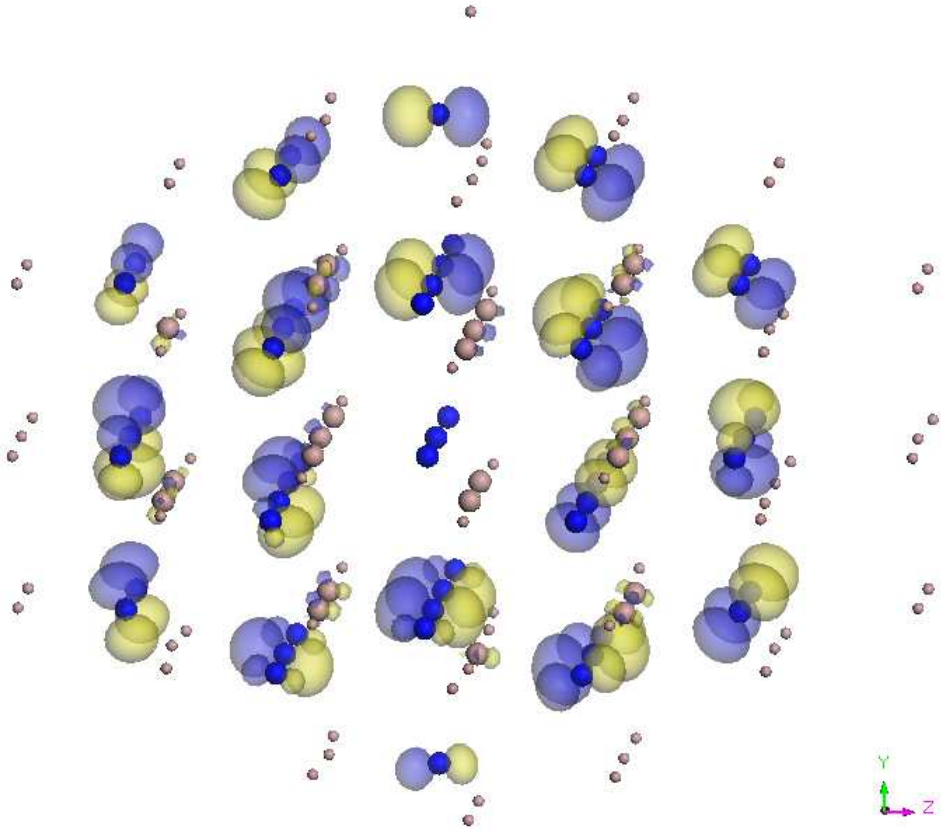


Figure 75: The HOMO of the 74-QM-atom cluster. The blue atoms are nitrogens and grey are Ga in Region I. Small grey Ga atoms are in Region II. The blue and yellow semi-transparent regions denote the positive and negative lobes of the HOMO respectively. Ions from Regions III-V are omitted.

ous studies of ZnO [16], convergence was improved by terminating the QM region with cations instead of anions and parameterising an anion ECP for Region II. A similar approach may prove fruitful in GaN but has not been attempted in the present work.

As has been noted before, the ionisation potential depends heavily on the polarity of the surface (see Section 3.8.2). The available value for the band gap at 0K is 3.505eV [9] and the electron affinity at 300K is 4.1eV [10] . This suggests an experimental ionisation potential of around 7.6eV, about 1eV higher than our result. We can also use the ZnO IP in conjunction

with the ZnO-GaN valence band offsets (the positions of the VBM of the two materials with respect to each other) to deduce the GaN IP. The IP value of ZnO is 7.71eV; the experiment and QM/MM calculation in [4] are in good agreement. The GaN-ZnO valence band offset has been variously reported as 0.7eV [17] (theoretical), 0.8eV [18] (experimental), 0.7±0.1eV ([19], experimental, polar surfaces) and 0.9± 0.1eV ([19], experimental, non-polar surfaces). Based on these data, the IP of GaN has been calculated as 6.7-7.1eV, less than the 7.6eV calculated from the GaN experimental band gaps and electron affinity, and closer to our results.

The lack of convergence of the IP to an asymptotic value makes it difficult to establish a definitive value for the the IP. However, it does not invalidate subsequent defect calculations using the present IP, provided that the appropriate value is used, i.e. the IP for the same QM-region size as that used in the defect calculation. The two calculations will then have a common reference level, the vacuum level for that QM-region size, which makes it possible to extract the relevant defect energies.

8.3.2 Ga vacancy

The formation energies of a Ga vacancy are summarised in Table 19. The results calculated using 42-QM-atom and 116-QM-atom clusters are shown. The results using the larger cluster differ by up to 0.4eV from the smaller cluster results, with the best agreement seen for the 1- triplet formation energy. While two cluster sizes are insufficient to determine the extent of the convergence of the formation energies, they do give a sense of the magnitude of errors that might be expected. In conjunction with the IP convergence studies in the previous section, we can conclude that the errors from the incomplete convergence with respect to the QM cluster size are of the order of a few tenths of an eV. In this case, the defects with charges 1- and more positive increase in formation energy with increasing QM cluster size while the more negative defects converge to lower formation energies.

The 42-QM-atom calculation predicts a relatively low formation energy for the Ga vacancy 1+ at the VBM, which is not observed for the calculation

Table 19: The formation energies (eV) of the gallium vacancy in the various charge states in the gallium-rich and nitrogen-rich limits at the VBM and CBM using 42-atom and 116-atom QM clusters and comparing with 96-atom LDA PBC calculations in [25] and [33], which uses the local spin density approximation (LDA) with up to 300-atom supercells. S = (spin) singlet, D = doublet, T=triplet, Q=quadruplet. The question marks refer to uncertainty in the spin state of the calculation in previous works.

Charge state	Ga-rich						N-rich	
	VBM		CBM		VBM		VBM	CBM
	# QM atoms in cluster:				PBC		# QM atoms:	
	42	116	42	116	Ref.[25]	Ref. [33]	42	42
1+ (S)	7.98	11.47	11.49	14.97			6.76	10.27
1+ (T)	9.39		12.90				8.17	11.68
0 (D)	10.00	10.43	10.00	10.43	9.06(?)	8.40(?)	8.79	8.79
0(Q)	9.95	10.33	9.95	10.33			8.73	8.73
1- (T)	12.61	12.74	9.11	9.24	9.31	8.83 (?)	11.39	7.89
2-	15.61	15.48	8.60	8.47	9.95	9.60	14.39	7.38
3-	19.02	18.62	8.51	8.11	11.05	10.67	17.80	7.28

with the 116 QM atoms. We offer two explanations, firstly the defect might be a resonance in the valence band. The energies of such resonances are not well defined. Secondly, the defect may be quite delocalised and the 42 atom QM cluster might not be sufficiently large to model it.

In the neutral charge state the spin quadruplet, i. e. a state in the which the three electrons in the dangling bonds are unpaired, is found to be slightly lower in energy than the doublet.

The formation energies as a function of Fermi level are shown in Fig. 76. Thermodynamic transitions are observed at much higher energies in the band gap than in the periodic calculations in [25] (see Table 20). If our predictions are correct, they might explain some of the DLTS results discussed in Section 8.4, which position defect levels 0.2-0.7eV below the conduction band.

The optical transition levels are shown in Table 21. The 1+ state is a resonance in the valence band. The 42-QM-atom optical transition levels are within about 0.2eV of the 116-QM-atom results where available.

Next, we plotted the HOMO of the 3- vacancy (Fig. 77), which can be seen

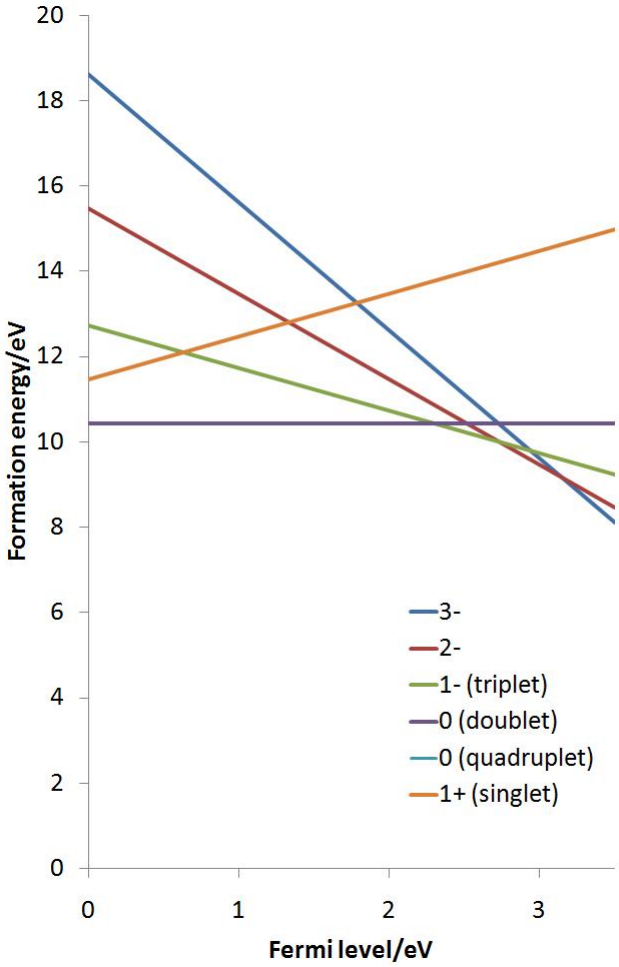


Figure 76: The formation energies of Ga vacancy in different charge states calculated with a 116 QM atom cluster. The thermodynamic transition levels are as follows: 0/1- at 2.31eV, 1-/2- at 2.73eV and 2-/3- at 3.15eV. The corresponding values for the 42-atom cluster are 2.66eV, 3.00eV and 3.41eV.

to consist of two p-type orbitals on the N ions adjacent to the vacancy. These correspond to the dangling bonds. The electronic structure of the 2- charge state is similar. The spin density of the neutral spin quadruplet is indicated in Fig. 78 and is distributed evenly over the four N neighbours adjacent to the vacancy as can be expected. The spin density of the spin triplet state of the charge 1- vacancy is shown in Fig. 79. The hole that has been filled relative to the neutral state is on the inequivalent N neighbour of the defect centre, along the c-axis (in orange) and the remaining holes are distributed

Table 20: The thermodynamic transition levels for the ionisation of a Ga vacancy, in eV above the VBM.

Transition	Present		Ref. [25]
	42 QM atoms	116 QM atoms	
3-/2-	3.41	3.15	1.10
2-/1-	3.00	2.73	0.64
1-(triplet)/0 (quadruplet)	2.66	2.31	0.25
0(quadruplet)/1+(singlet)	1.97	×	×

Table 21: The optical transition levels for the ionisation of a Ga vacancy, in eV below the CBM. RES denotes resonances.

Transition	116 QM atoms	42 QM atoms	
	B97-2	B97-2	SVWN
3-/2-	0.48	0.34	0.56
2-/1-	1.03	0.80	
1-(triplet)/0 (quadruplet)		1.17	
0(quadruplet)/1+(singlet)		5.13 (RES)	

evenly over the three in-plane N nearest neighbours (pink). Three lobes of spin density are formed on each of the three equivalent N nearest neighbours.

Table 19 on p. 208 also shows that our values of formation energies are significantly larger than those calculated with periodic boundary conditions, especially for the more negative charge states. Under Ga-rich conditions, [28] finds the formation energy of the Ga vacancy in the charge state 3- at the CBM to be around 1.5eV in contrast to our calculations of about 8.5eV. There is better agreement for the less negatively charged Ga vacancies. A summary of various calculations of the neutral Ga charge state is shown in Table 22. Our calculations are at the higher end of the spectrum.

We investigated two possible reasons for the discrepancy between our and PBC results: the functionals and basis sets employed. There are, of course, multiple differences between our method and the PBC, which are discussed in more detail in Section 8.4.2.

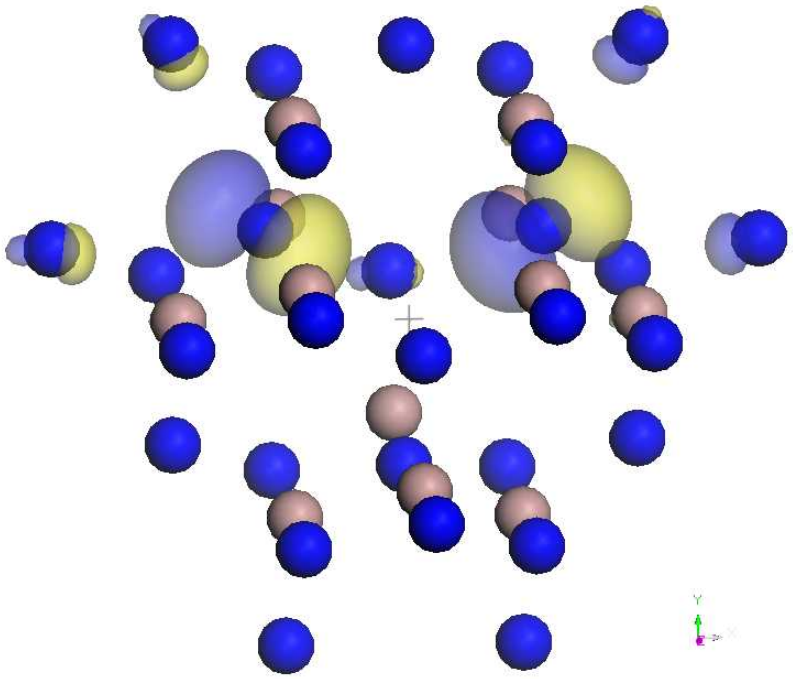


Figure 77: The HOMO of the charge 3- Ga vacancy. The defect centre is indicated by a cross. The 0.03e/Å^3 isosurface is shown. The +/- lobes of the orbital are depicted in semi-transparent blue and yellow.

Table 22: The formation energies of the neutral Ga vacancy from a range of theoretical studies under N-rich conditions (eV), from [35] and references therein. [35] used self-consistent-charge density-functional-tight-binding (SCC-DFTB) whereas the other works cited use DFT for defect formation energy calculations.

Present	[25]	[35]	[36]	[32]	[37]	[38]	[39]
8.73	7.84	7.83	6.14	6.8	6.3	6.3	6.7

Our calculations use the B97-2 hybrid functional whereas the PBC calculations in [28] employ the LDA. We investigated the effect the change in the functional would have on our calculations and used the LDA SVWN functional [21] as implemented in Chemshell. The results are summarised in Table 23. The differences between the SVWN and B97-2 are of the order of a

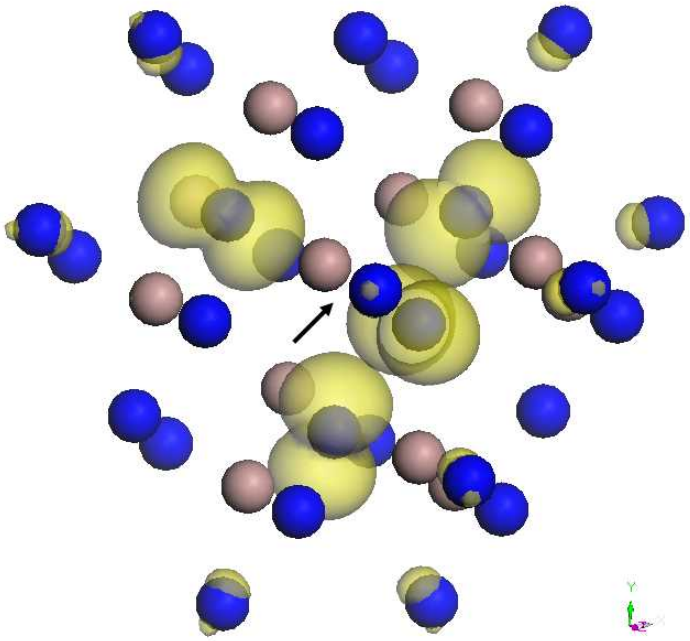


Figure 78: The spin density of the neutral Ga vacancy. The defect centre is shown by an arrow. The 0.005e/Å^3 isosurface is indicated. The spin density is depicted in semi-transparent yellow.

few 0.1eV, insufficient to explain the large discrepancies between the present calculations and the PBC results in [28]. The optical transition levels, shown in Table 21 on p. 210 are also in fairly good agreement. A further issue is the spin state of the electronic configuration calculated with PBC, which is unclear from the literature. A more detailed discussion of the discrepancies is provided in Section 8.4.

In our most recent calculations we re-considered our basis set. We re-introduced the diffuse l function originally taken out to increase the speed of the calculations[13]. For the case of the 42-atom calculations, this expanded basis set resulted in a reduction of the formation energy for the two test cases with charge 1- and 3-, of around 2.5eV almost exclusively due to the changed energy of the free Ga atom.

Furthermore, we recalculated the 116-QM-atom cluster Ga vacancy forma-

[13] l Ga 1.0000000 0.0746100 1.0000000

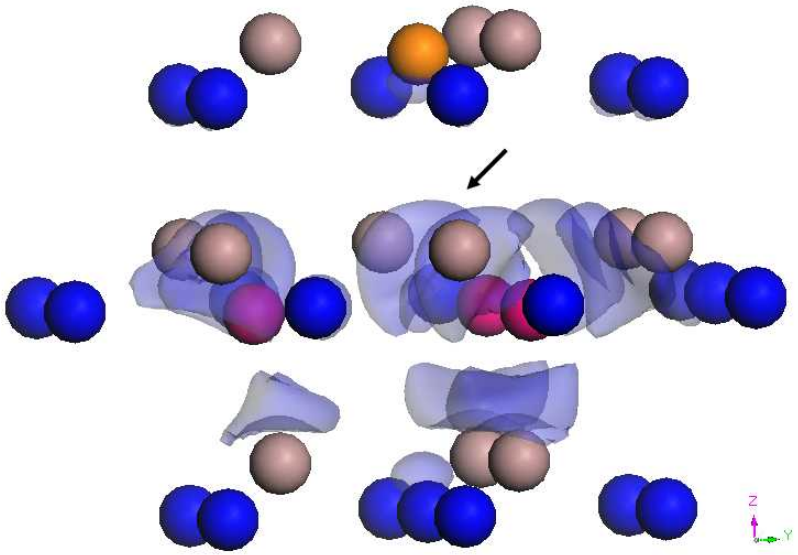


Figure 79: The spin density of the Ga vacancy in the charge state 1-. The defect centre is shown by an arrow. The 0.005e/Å^3 isosurface is indicated. The three equivalent N are in pink and the inequivalent one is in orange. The spin density is depicted in semi-transparent blue.

Table 23: Formation energies of the gallium vacancy in two charge states in the gallium-rich limit at the VBM calculated with different energy functionals and a 42-QM-atom cluster.

Charge state	B97-2	SVWN
3-	19.02	18.47
2-	15.61	15.56

tion energy by placing a VTZ basis set due to Peterson (see Appendix, p. 269 and [46]) on the vacancy defect centre. The same basis set and the associated ECP [47] were used to calculate the energy of the free Ga atom. This set-up resulted in an even more dramatic drop in formation energy, of about 3.5eV. We conclude that the abbreviated basis set is insufficient to model the Ga vacancy and the free Ga atom. We recommend further investigations using

Table 24: The geometry of the Ga vacancy - distances from the defect centre to the nearest N neighbours in Å.

Charge state	Distances to nearest neighbours
2-	2.515, 2.398, 2.301, 2.292
3-	2.546, 2.381 (×2), 2.363

the expanded basis set and the 116 atom cluster for the full range of charge states. These calculations will require considerable computational resources.

8.3.3 Ga interstitial

The formation energies for the Ga interstitial are shown Fig. 80. We note that the 3+ and 2+ charge states have exothermic formation energies at the VBM, the same feature encountered with the nitrogen vacancy later in this chapter and discussed as a possible reason for the difficulty in p-doping GaN in Section 8.4. For the 3+ charge state, the energy released is more than 4eV. Table 26 shows that the 4+ charge state is a resonance in the valence band as the 3+/4+ transition with respect to the VBM is positive. The neutral charge state is a resonance in the conduction band.
The formation energy results are also tabulated in Table 25 and compared to PBC calculations with 96-atom supercells using the LDA approximation in [25], which predict an endothermic formation energy at the VBM for the 3+ and 2+ charge states. There is a negative offset in our values of about 6.2-6.5eV relative to the PBC results for all the positive charge states
Recently, we carried out calculations with the Peterson basis set on the interstitial Ga atom and added the diffuse l function to the other Ga atoms in the QM cluster as described in Section 8.3.2 on p. 210 for the Ga vacancy calculation. The formal charge state formation energies were determined to be 8.64 eV for the 74-QM-atom cluster and 9.27eV for the 116-QM-atom cluster at the CBM in the Ga-rich limit, 2.39eV and 3.02 eV respectively above the values calculated with the abbreviated basis set, which brings the val-

ues closer to the PBC results. Further calculations on the remaining charge states for the 116-QM-atom set-up are recommended for future analysis.

The trends in the relative stability of the different charge states using the original set-up without the diffuse l function are similar to those observed in periodic DFT calculations (see Table 25). The 3+ charge state is the most thermodynamically stable charge state near the VBM and the 1+ charge state at the CBM, and 2+ the most stable species at only a small range of Fermi levels. The 3+/2+ thermodynamic transition occurs at at a Fermi level of 2.56eV, 2+/1+ at 2.69eV. Periodic DFT calculations in [25] predict

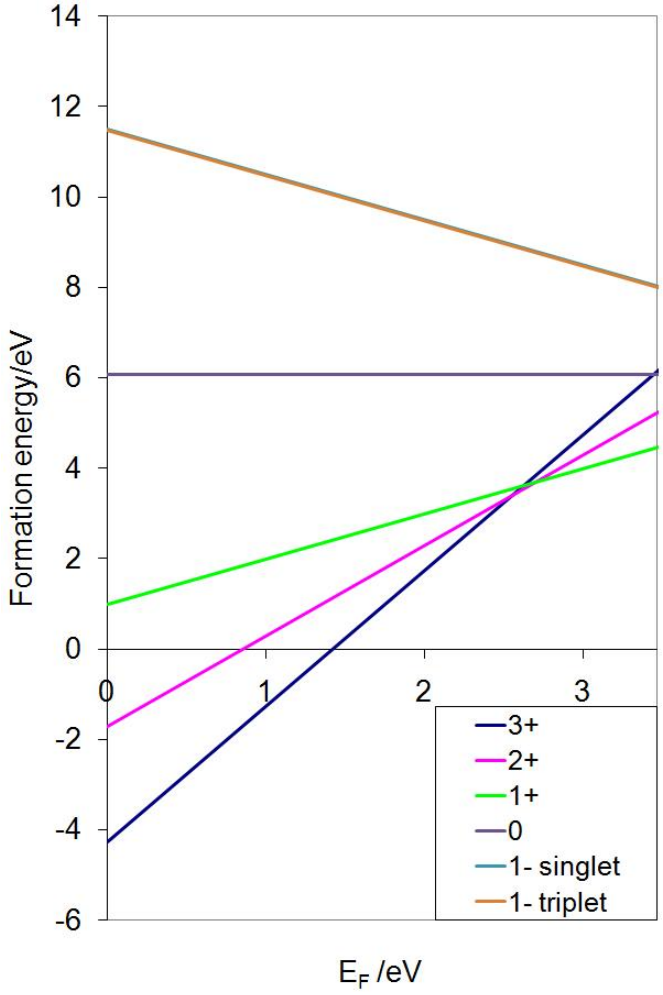


Figure 80: The formation energy of a Ga interstitial.

Table 25: The formation energies of the gallium interstitial in various charge states. PBC results from [25] are also quoted.

	Ga-rich			N-rich	
Charge state	VBM	CBM	VBM [25]	VBM	CBM
3+	-4.27	6.25	2.34	-3.05	7.47
2+	-1.71	5.30	4.89	-0.49	6.52
1+	0.98	4.48	7.28	2.20	5.71
0	6.08	6.08		7.30	7.30
1- (triplet)	11.47	7.96		12.69	9.19
1- (singlet)	11.49	7.99		12.72	9.21

Table 26: The optical transition levels (eV) for the ionisation of a Ga interstitial and its electron affinity. RES denotes resonance.

Ionisation energy w.r.t			Electron affinity w.r.t		
Transition	VBM	CBM	Transition	VBM	CBM
0/1+	-4.92	-1.42 (RES)	1+/0	5.26	1.75
1+/2+	-1.63	1.88	2+/1+	3.43	-0.07
2+/3+	-1.30	2.21	3+/2+	3.11	-0.39
3+/4+	1.39 (RES)	4.90			

the 3+/1+ transition (i.e. a negative-U defect) at 2.47eV, in good agreement with our results in terms of the value although we predict the charge 2+ state to be the thermodynamically most stable species for a range of Fermi levels of about 0.1eV. Again, parallels can be drawn with the observations for the N vacancy later in the chapter. The neutral charge state is an unstable resonance in the conduction band (see Table 26).

The equilibrium geometry of the Ga interstitial is an octahedral arrangement with a reduced symmetry. The interstitial lies in the hexagonal channel of the wurtzite structure and can be thought of as sandwiched between two planes of cations with normals along the c-direction. Similarly, it lies between two layers of anions. There are three nearest N neighbours in one of the planes and three in the other plane. Two out of three nearest neighbour N ions in each plane are equivalent and have equal distances r_{equiv} to the inter-

Table 27: The geometries of the Ga interstitial in different charge states. The distances from the interstitial to the nearest neighbours X are shown.

Charge	X	Interionic distance from interstitial Å to Two equivalent ions	One inequivalent ion
3+	N (plane 1)	1.933	1.908
	N (plane 2)	2.610	2.686
	Ga (plane 1)	2.574	2.569
	Ga (plane 2)	2.486	2.504
2+	N (plane 1)	2.009	1.907
	N (plane 2)	2.781	2.916
	Ga (plane 1)	2.528	2.615
	Ga (plane 2)	2.522	2.441
1+	N (plane 1)	2.105	1.931
	N (plane 2)	2.899	3.096
	Ga (plane 1)	2.478	2.718
	Ga (plane 2)	2.612	2.282
0	N (plane 1)	2.117	1.934
	N (plane 2)	2.925	3.146
	Ga (plane 1)	2.477	2.720
	Ga(plane 2)	2.638	2.275

stitial, while the third one, r_{inequiv}, is different, $r_{\text{equiv}} \neq r_{\text{inequiv}}$. The situation is analogous for the Ga nearest neighbours, with two planes containing three nearest neighbours each, two equivalent and one inequivalent. The interionic distances are summarised in Table 27. We note that the geometry distortion from a perfect octahedral arrangement, i.e. the difference between r_{equiv} and r_{inequiv}, becomes more pronounced as the charge state becomes less positive. Compared with the results of the MM studies in Table 9, the interstitial - N bond lengths for the charge state 3+ shorter by about 0.1 Å. The Ga-Ga distances calculated with the interatomic potential are shorter (for one set of MM distances) and longer (for the other set of MM) distances than the QM/MM values also by around 0.1 Å.

Because of the octahedral arrangement, atomic orbitals are a good starting point for understanding the electronic structure. The Ga^{3+} interstitial can be expected to be a closed shell argon-type configuration. An extra electron would be placed into the s-type orbital, in a similar fashion to a Ga^{3+} to

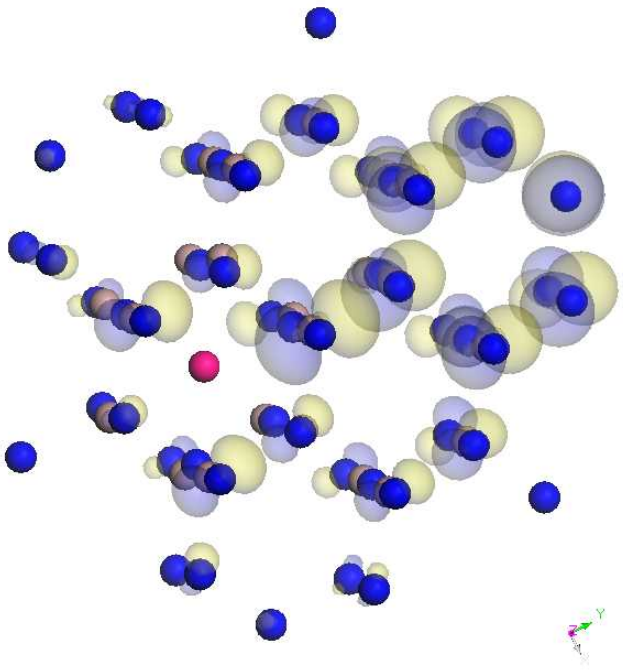


Figure 81: The HOMO of the Ga interstitial 3+, isosurface 0.03eV/Å^3. The electron density mostly resides in p-like orbitals on the cluster N and are concentrated on N on the edge of the cluster, far away from the interstitial.

Ga^{2+} free ion transition.

The HOMO of the 3+ Ga interstitial is shown in Fig. 81 and contains the p-type orbitals on the Ns in the cluster as expected. The 3d electrons of the Ga are lower in energy.

The 2+ and 1+ charge states have more complicated electronic structures with a single lobe positioned on the Ga interstitial. In a free Ga ion with charge 1+ or 2+ the HOMO can be expected to be a spherical s-type orbital. In the crystal environment, this orbital is distorted as shown in Fig. 82. The xy-plane view is shown in Fig. 83.

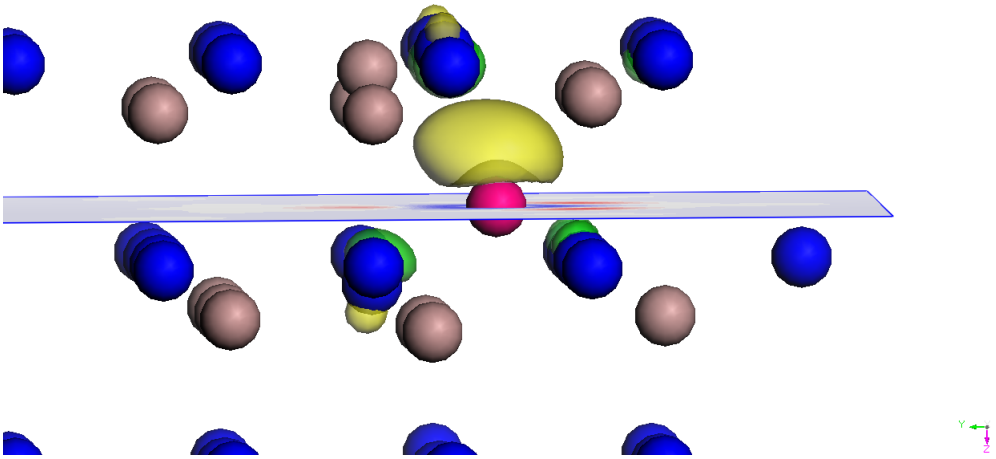


Figure 82: The HOMO of the Ga interstitial 2+ (yellow and green lobes). The xy-plane is also indicated. The electron density in this plane is shown in Fig. 83. The interstitial is in pink, Ga in grey and N in blue.

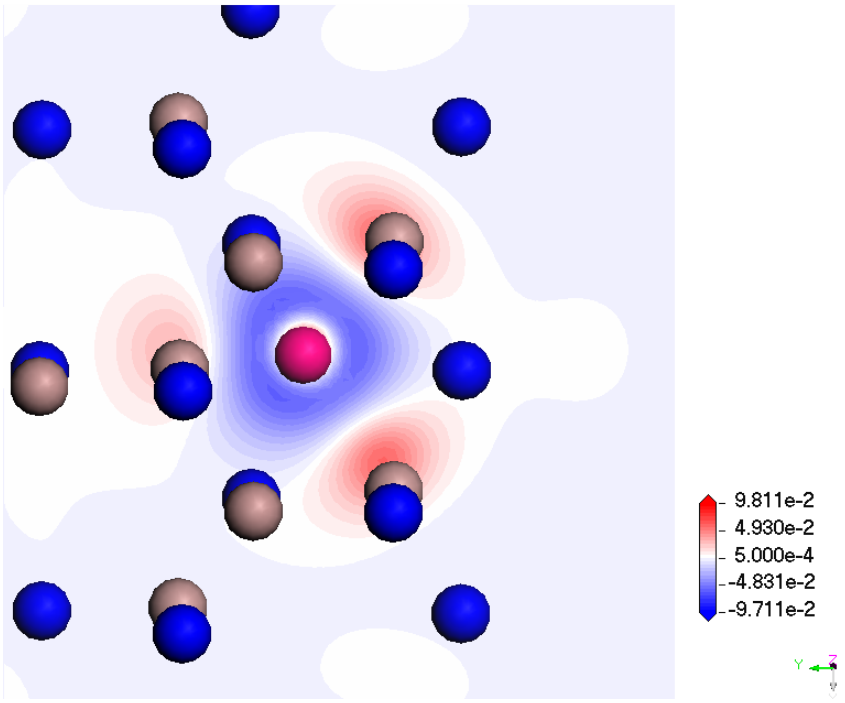


Figure 83: The HOMO of the Ga interstitial 2+, xy-plane view.

8.3.4 N interstitial

A number of configurations are in principle possible for the N interstitial. It could assume octahedral or tetrahedral sites as has been described for Ga interstitials (see Fig. 45). Most studies, however, report a split interstitial configuration, apart from [22], which discusses other geometries, in particular for the more negative charge states. In the 3- charge state, the N interstitial might feasibly assume a channel-centre configuration.

The split interstitial configuration is the lowest energy geometry as calculated with the MM model in Section 6.2, albeit with a large bond length so the interstitial is found close to the centre of the hexagonal channel and the lattice N moves off its lattice site. The split configuration was used as the starting point of the present calculation to speed up the expensive optimisation process. In the future, other starting geometries may be explored. Some of the configurations in [22] would be possible candidates.

The split interstitial is a N_2-like configuration and therefore the molecular orbital energy levels of this molecule are a useful starting point for the consideration of the interstitial. The electronic configuration of N_2, as shown in Fig.84, corresponds to the N interstitial in the charge state 3+ since the molecule is neutral and the lattice nitrogen that the interstitial pairs up with has a charge of 3-. On adding two more electrons, to form the 1+ charge state, one would expect the electrons to occupy the N_2 π^* antibonding orbitals. The spins can be expected to be unpaired (spin triplet) although we also calculate a paired configuration (spin singlet) and compare its energy with the triplet. Adding two extra electrons to form the 1- charge state results in a closed shell configuration.

The expectation of a spin triplet for the 1+ charge state is indeed borne out by our calculations: the optimised charge 1+ spin triplet state is approximately 1eV lower in energy than the spin singlet.

The formation energies for charge states varying from 3+ to 2- are shown in Table 28, in the nitrogen-rich and gallium-rich limits (see Section 3.8.1). The dependence of the formation energy on the Fermi level, based on Eq. 63, is shown in Fig. 85.

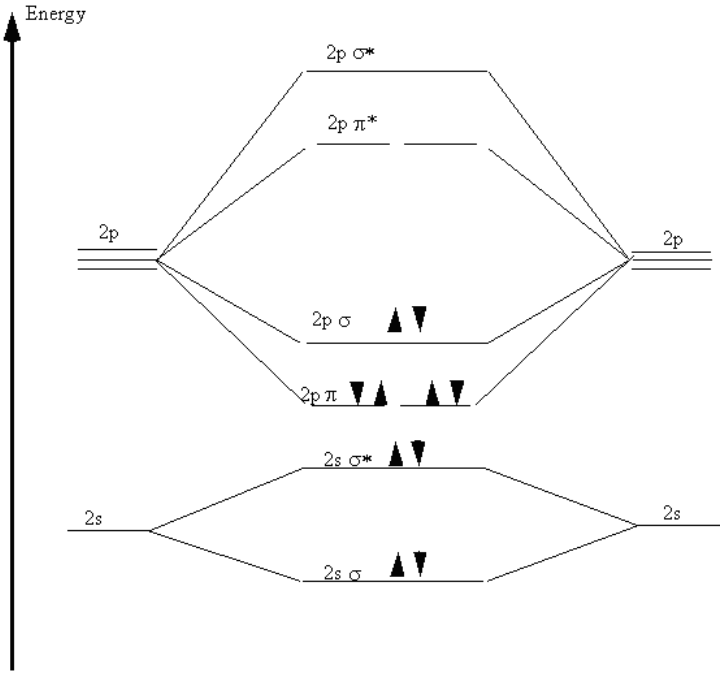


Figure 84: The energy levels of a nitrogen molecule. From [23]

The 3+ charge state is a resonance in the valence band and the 2- charge state a resonance in the conduction band, according to the optical transition levels shown in Table 29 on p. 223 . The formation energy of the 3- species was not calculated on account of the current method being unreliable in studying the delocalised conduction band states. The 3+ charge state is also expected to be a resonance in the valence band. Future work may include the calculation of the 3+/2+ electron affinity to determine the stability of the resonance. A negative value of electron affinity with respect to the VBM implies that the charge state spontaneously accepts an electron from the VBM and implies an extremely short-lived state.

The present formation energies in Table 28 are compared with those obtained with periodic boundary conditions DFT in the local density approximation, with 96-atom supercells in [25]. Our values for the formation energies are generally lower at the VBM for positively charged and neutral defects, by

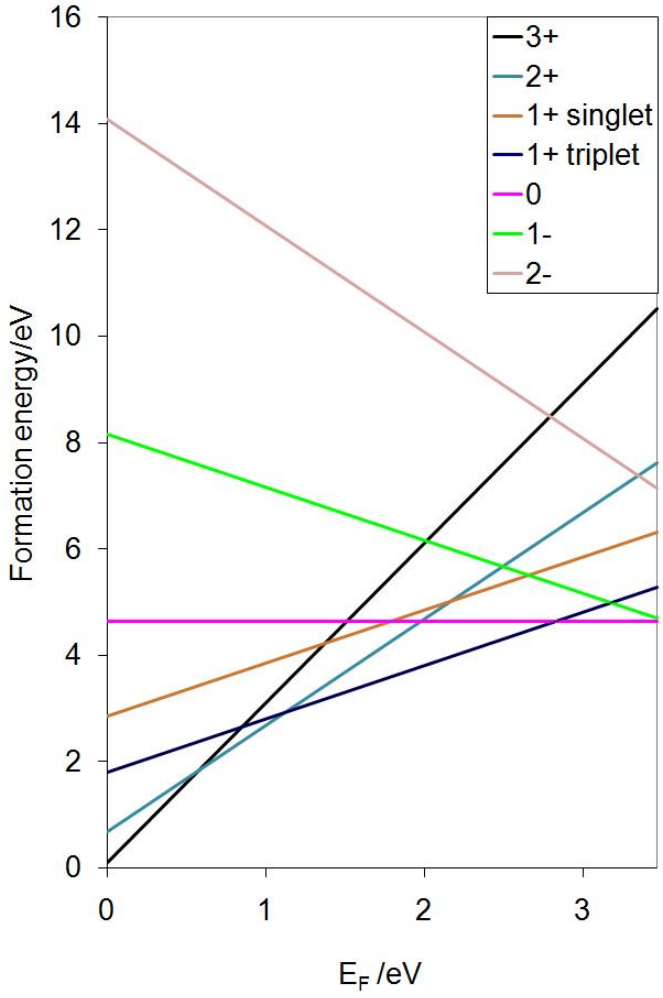


Figure 85: The formation energies of nitrogen interstitial as a function of Fermi energy for different charge states in the nitrogen-rich limit. 0 on the x-axis denotes the VBM. The CBM is at 3.505eV (i.e. one band gap energy above the VBM). The 2+/1+(triplet) thermodynamic transition, the Fermi level at which 1+ and 2+ charge state formation energies cross, is 1.13eV above the VBM and the 1+(triplet)/0 thermodynamic transition occurs at 2.82eV above the VBM. The 2+/3+ transition is 0.58 above the VBM.

about 0.5-2eV. The negatively charged states in the present work have formation energies about 1-2eV higher.

Reference [25] does not report optical transition levels. The thermodynamic defect levels were calculated from their VBM formation energies, assuming

Table 28: The formation energies of the nitrogen interstitial in the various charge states at the VBM and CBM. The fourth results column gives the data from [25], which used periodic boundary conditions within the local density approximation and 96-atom supercells. It is unclear what the spin state of the N interstitial in the 1+ charge state in [25] is.

Charge state	Formation energy (eV)				
	N-rich conditions		Ga-rich conditions		
	VBM	CBM	VBM	VBM [25]	CBM
3+	0.10	10.61	1.32	3.19	11.84
2+	0.67	7.68	1.90	3.93	8.91
1+ (triplet)	1.80	5.31	3.03	4.83 (triplet?)	6.53
1+ (singlet)	2.85	6.35	4.07		7.57
0	4.63	4.63	5.85	6.31	5.85
1-	8.16	4.66	9.38	8.31	5.88
2-	14.08	7.07	15.30	12.90	8.29

Table 29: The optical transition levels in eV for the nitrogen interstitial. The two types of defect level relate to the ionisation of a defect and its electron affinity. T refers to the spin triplet state of the 1+ interstitial. Negative values refer to states *above* the relevant level, i.e. VBM or CBM. States above the CBM are resonances (RES) which autoionise to the bottom of the conduction band.

Ionisation energy (eV)			Electron affinity (eV)		
Transition	VBM	CBM	Transition	VBM	CBM
2-/1-	-6.03	-2.53 (RES)			
1-/0	-2.83	0.67	0/1-	4.43	0.92
0/1+(T)	-1.77	1.73	1+(T)/0	3.76	0.25
1+(T)/2+	-0.04	3.47	2+/1+(T)	2.04	-1.47
2+/3+	0.50 (RES)	4.00			

a band gap of 3.505eV. The 2+/3+ and 2+/1+ transition levels are in reasonably good agreement, 0.2eV apart although it is not clear whether their 1+ calculation is for a spin triplet or singlet. The 0/1+ transition shows less good agreement. We do not predict a 1-/0 thermodynamic transition for the nitrogen interstitial, unlike [25]. At the CBM, our charge states 0 and 1-

Table 30: The thermodynamic transition levels with respect to the VBM, deduced from the crossings in Fig. 85. Values given above the VBM (positive) and below the CBM (negative).

Transition	VBM		CBM
	Present	[25]	Present
2-/1-	×	×	
1-/0	×	2.00	
0/1+(triplet)	2.82	1.48 (triplet?)	-0.69
1+(triplet)/2+	1.13	0.90 (triplet?)	-2.38
2+/3+	0.58	0.74	-2.93

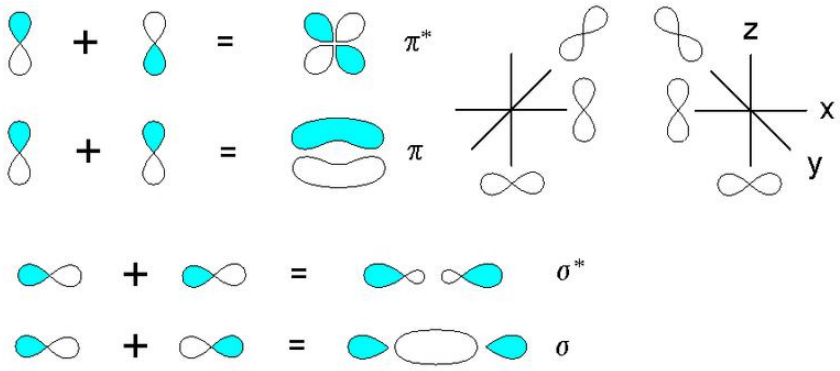


Figure 86: The σ and π bonding and antibonding orbitals. From [24].

are degenerate within the accuracy of the calculation. For a summary of the thermodynamic transitions see Table 30.

One would expect the strongly localised, deep defect levels to be most suitable for QM/MM treatment. The extent of the delocalisation is best assessed directly by plotting the HOMOs. The electron density of the N split interstitial with charge 0 is shown in Fig. 87. The shape of the orbital is consistent with an N_2 π^* (anti-bonding) orbital (see Fig. 86), in agreement with Fig. 84, which predicts the electronic structure of the split interstitial on the basis of the corresponding levels in a N_2 molecule. Adding three electrons to the configuration in Fig. 84 results in charge state 0. The HOMO is then indeed an N_2 π^* orbital.

As a measure of the delocalisation of the electrons in the 1+, 0 and 1-

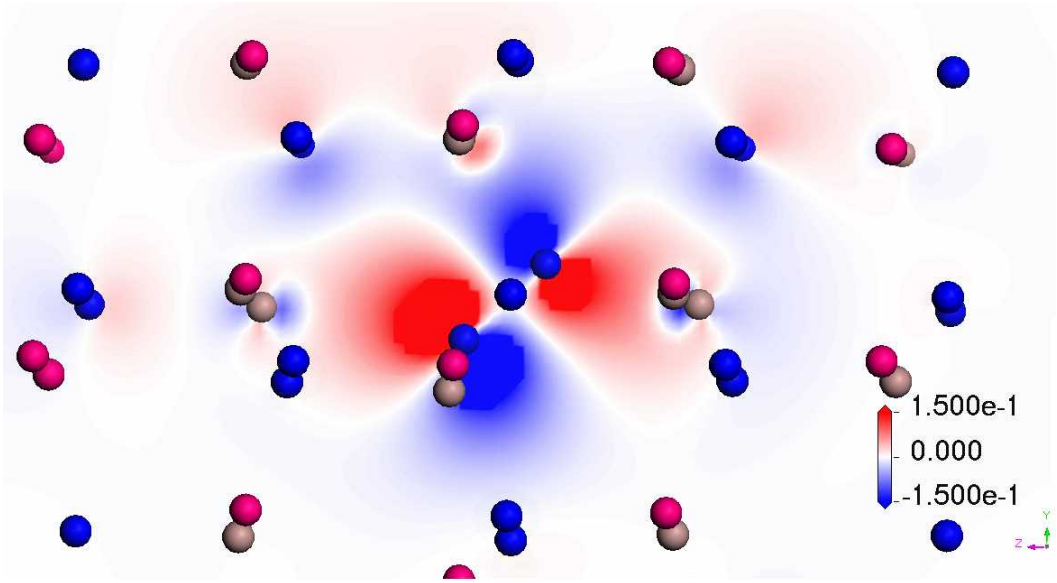


Figure 87: The HOMO in the plane of the split interstitials for N interstitial charge 0.

Table 31: The volumes enclosed by the 0.05e/Å^3 isosurface for N split interstitials.

Charge state	Volume enclosed (Å^3)
1-	4.941
0	5.082
1+	4.875

charge states the volumes enclosed by the 0.05e/Å^3 isosurface are shown in Table 31. The volume increases from 1+ to 0 charge state, consistent with an increased electronic repulsion. The 1- charge state becomes slightly more localised by this measure.

Finally, the defect geometries, in this instance the distance between the interstitial ion and the lattice nitrogen with which it forms a split formation, are presented in Table 32. A comparison is made with the results in [32], where periodic boundary conditions were used for 32-atom supercells, which is a relatively small supercell size for defects with strong relaxations. Nonetheless, a good agreement is shown with our calculations unlike in the

Table 32: The geometries for the nitrogen interstitial.

Charge level	N-N split distance (Å)	% change relative to N_2 bond length, 1.1Å	
		Present	[32]
1+	1.258	14	14
0	1.350	23	22
1-	1.458	33	30

case of the energy calculations. The geometry of the split interstitial is shown in Fig. 88. As has been proposed in [22], the N interstitial might assume a low energy configuration in the centre of the hexagonal channel, especially in the 3- state, which might lower its formation energy. This configuration has been found by our MM formal charge calculations (see Fig.48), which, being computationally less expensive, allowed for a more extensive configuration search. Our QM/MM search might have found a local, rather than global, minimum because the final geometries of a particular charge state were used as the starting geometries for the next more negative charge state. Such a procedure might not find the correct global minimum if there is a large difference between the geometries of the lowest energy configurations for the two charge states. The investigation of other defect configurations using the QM/MM approach would be an interesting subject for future study.

8.3.5 N vacancy

We now turn our attention to the energetics of nitrogen vacancies. The formation energies in the N-rich limit as a function of Fermi level are shown in Fig. 89. The formation energies at the VBM and CBM in the N-rich and Ga-rich limits are given in Table 33.

The formation energy of the 3+ and 2+ level at the VBM becomes negative, implying that the formation is spontaneous. The 3+ state is a resonance in the valence band. The 3+ state is possibly a resonance in the valence band although the 2+/3+ optical transition is only 0.06eV below the VBM (see

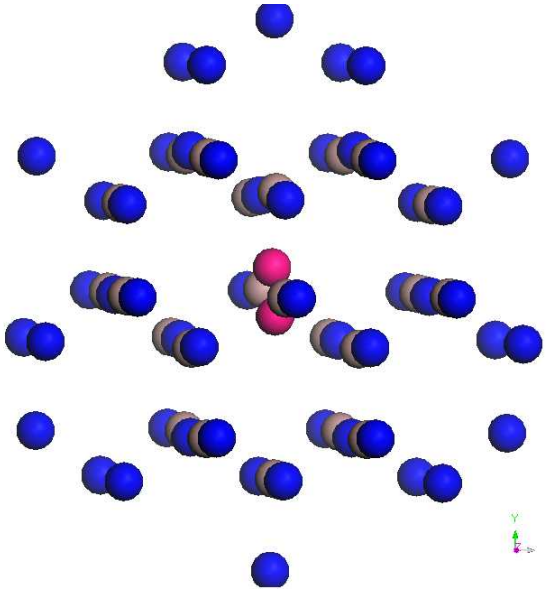


Figure 88: The geometry of the 1- split N interstitial (view along the z-axis). The interstitial and the lattice N with which it forms a bond are shown in pink.

Table 33: The formation energies of the nitrogen vacancy in the various charge states at the VBM and CBM. Comparing with the results at the VBM [25] using 96-atom supercells in periodic boundary condition DFT calculations in the LDA approximation using the non-linear core correction (nlcc) for the description of 3d electrons. [33] uses the PAW method for treating 3d electrons.

Charge state	Formation energy (eV)					
	N-rich		Ga-rich			
	VBM	CBM	VBM	CBM	VBM [25]	VBM [33]
3+	-2.31	8.21	-3.53	6.99	-1.08	0.89
2+	-1.19	5.81	-2.41	4.60		0.95
1+ (singlet)	0.015	3.52	-1.21	2.30	0.10	0.82
1+ (triplet)	2.60	6.10	1.38	4.88		
0 (doublet)	4.06	4.06	2.84	2.84		3.16
1- (singlet)	8.15	4.64	6.93	3.42		5.00
2-	13.50	6.49	12.28	5.27		8.45
3-						10.59

Table 34). In Section 8.4 we relate the negative formation energies to hole

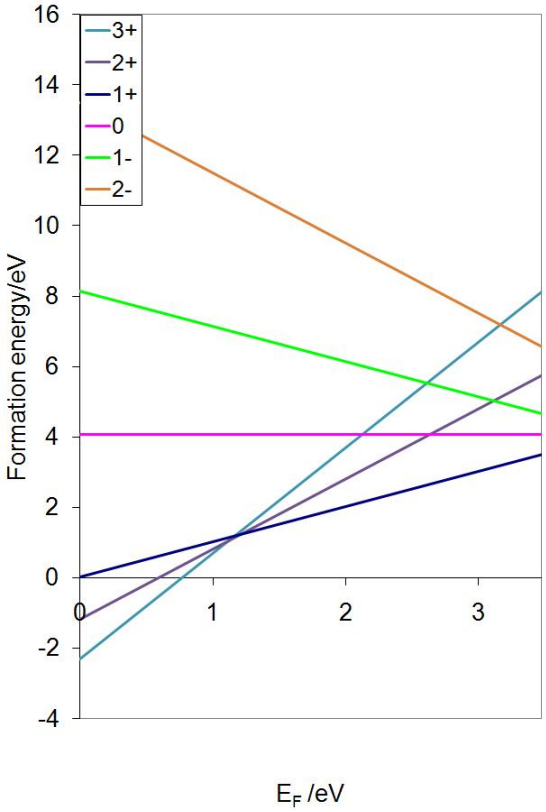


Figure 89: The formation energies of the nitrogen vacancy in the various charge states as a function of Fermi level in the nitrogen-rich limit. A thermodynamic 3+/2+ transition is observed at 1.12eV and 2+/1+ at 1.21eV.

Table 34: The optical transition levels (eV) for the ionisation of a nitrogen vacancy and its electron affinity

Ionisation energy w.r.t			Electron affinity w.r.t		
Transition	VBM	CBM	Transition	VBM	CBM
2-/1-	-4.83	-1.33			
1-/0	-3.33	0.18	0/1-	5.04	1.53
0/1+	-3.08	0.42	1+/0	4.74	1.24
1+/2+	-0.13	3.38	2+/1+	1.66	-1.85
2+/3+	0.06	3.57			
3+/4+	1.51	5.02			

Table 35: The formation energies of the neutral N vacancy from a range of theoretical studies under N-rich conditions (eV). From [35] and references therein. [35] used self-consistent-charge density-functional-tight-binding (SCC-DFTB) whereas the other works cited use DFT for defect formation energy calculations.

Present	[35]	[36]	[32]	[37]	[38]	[39]
4.06	3.65	4.81	1.2	4.6	5.0	2.8

compensation behaviour of GaN and the difficulty of p-doping the material. Periodic DFT studies in [25] (see Table 33) using 96-atom supercells also report negative formation energies at the VBM.

Other DFT calculations have predicted formation energies that are quite different from ours, often by several eV. These include periodic calculations in [25] performed using 96-atom supercells and the non-linear core correction, whose results are presented in Table 33. The results from [33], which used the projector augmented wave method (PAW) are also shown. The best agreement with these results is for the neutral charge state.

A number of other results for the neutral state are listed in Table 35. There is significant variation among the results calculated with periodic DFT methods using a number of approximations. The formation energies range from 1.2 to 5.0eV. Our result is within this range of variation.

In Fig. 89 it can be seen a thermodynamic 3+/2+ transition occurs at 1.12eV above the VBM and 2+/1+ at 1.21eV. The PBC calculations within the LDA approximation and with the non-linear core correction [40] in [25] predict a 3+/1+ transition, suggesting that the nitrogen vacancy is a negative-U defect (see Section 3.8.3), at about 0.59eV above the VBM. Both our work and [25] predict that the 1+ charge state of the N vacancy is the most thermodynamically stable at the CBM, which is in contrast to the results in [33], with charge 1- being the thermodynamically stable state at the CBM and formation energies comparable to those of gallium vacancies. This latter work used the local spin density method and the projector augmented wave method (PAW) to treat the core electrons including the 3d shell of Ga.

The nitrogen vacancy is expected to be a potential well for electrons because the four gallium ions surrounding the original nitrogen site are positively charged and, the defect site has a deficiency of electrons.
The molecular orbital picture can be used as a starting point for rationalising the electronic structure. In the case of the N vacancy in the charge state 3+, all the orbitals pointing from the nearest neighbour Ga ions to the defect site are empty. Adding an extra electron to the 3+ charge state will result in an open-shell configuration and an electron might be expected to be delocalised relatively evenly over the four neighbouring gallium ions, owing to the large size of the gallium ion: the orbitals available for the electron from the defect site are relatively far from the gallium core and quite delocalised. One would therefore expect a relatively spherical distribution of electrons around the defect site. The Jahn-Teller effect would be less likely to occur as it is usually associated with localised electrons. However, our calculations predict the 3+ charge state to be a resonance in the valence band. A near-spherical HOMO is also predicted for the 1+ charge state, a closed shell configuration. The electronic structure of more highly charged defects is more difficult to predict but more localisation would be expected.
The calculated electronic structure of the N vacancy, i.e. the highest occupied molecular orbital (HOMO), in the charge state 1+ is shown in Fig. 90 (plane view) and Fig. 91 (3-D view). As predicted the distribution is approximately spherical. It is quite a delocalised state, with significant charge density residing on the neighbouring nitrogen ions.
The charge states of 0 and 1- have a more unusual charge distribution. The spin density (see Section 2.2.3 on p. 56) of the neutrally charged state is shown in Fig. 92. The spatial distribution is equivalent to the HOMO. There are two orthogonal lobes, each connecting two nearest Ga neighbours of the defect centre (in pink). The lengths of the bonds between the Ga atoms connected by the HOMO lobes are 2.894 and 3.002 Å.
The 1- charge state has the same electronic structure with shorter Ga-Ga bond lengths of 2.683 and 2.644 Å. The xy- plane projection of the charge density is shown in Fig. 93.
This electronic structure is similar to that suggested in [32]. They interpret

the geometry as strongly overlapping Ga dangling bonds forming a metallic-like bond and suggest that the reason for the strong overlap is that in bulk GaN the Ga-Ga distances are similar to those in Ga metal. They calculate the electronic structure of the arsenic vacancy in GaAs, with a 20% larger Ga-Ga distance and do not observe similar behaviour. The calculations in [33] and the empirical tight-binding calculations in [42] also note the strong Ga-Ga interaction surrounding the N vacancy.

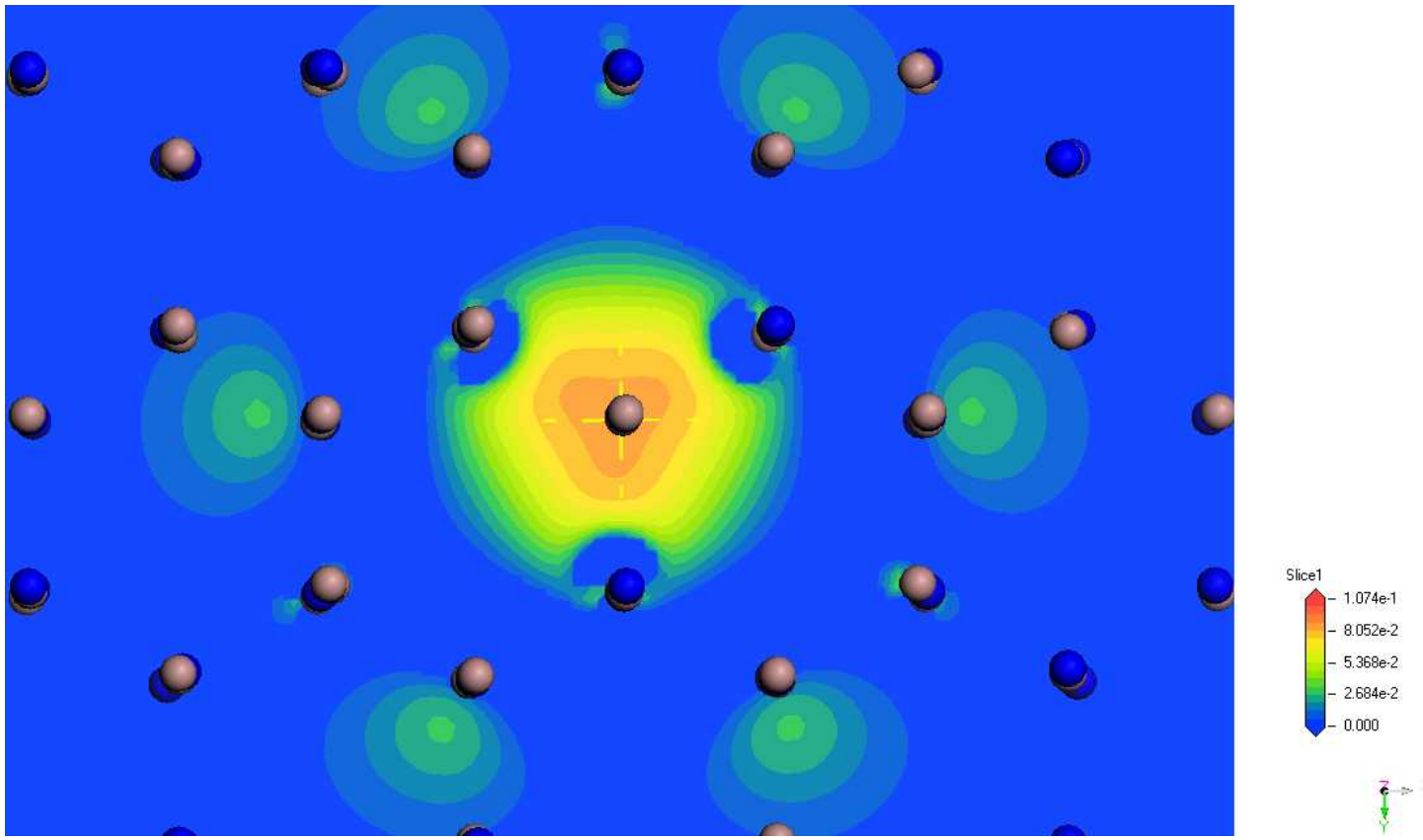


Figure 90: An xy-plane view of the HOMO of the N vacancy in the 1+ charge state. Ga atoms in grey, N atoms in blue and the defect centre is in pink.

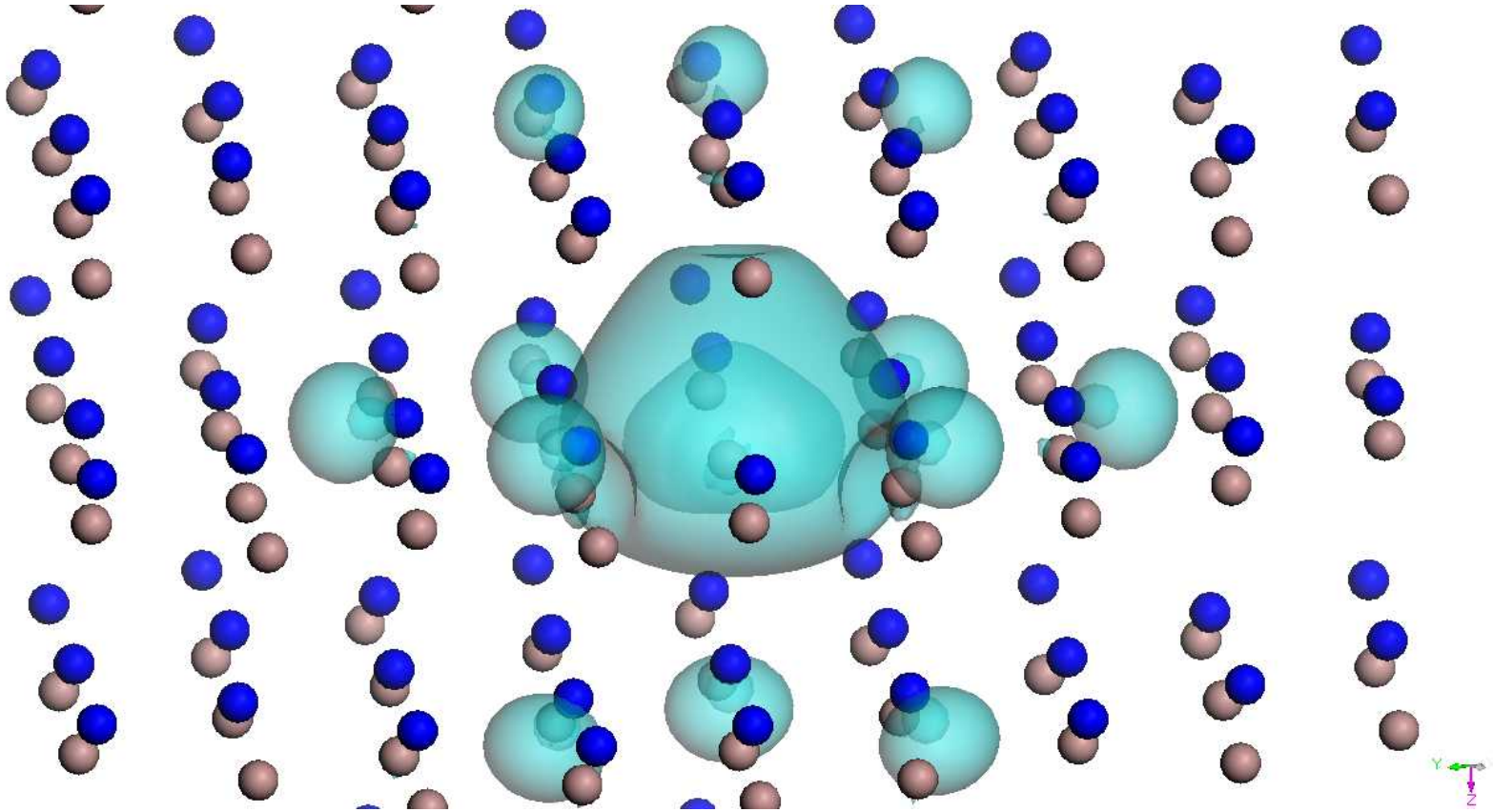


Figure 91: The HOMO of the N vacancy in the 1+ charge state. Ga atoms in grey, N atoms in blue. The isosurfaces shown are those of charge densities 0.03 and 0.07 e/Å^3.

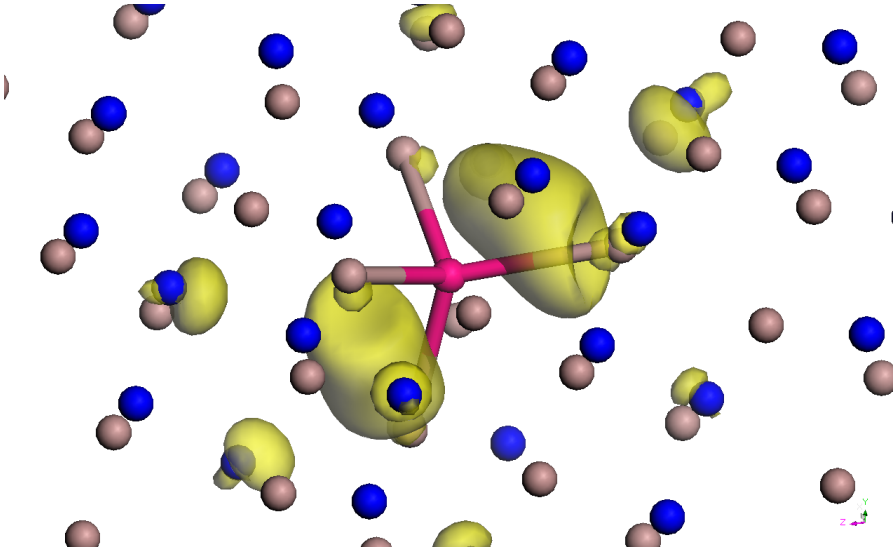


Figure 92: The spin density (yellow) of the neutrally charged N vacancy. Ga is grey and N is blue.

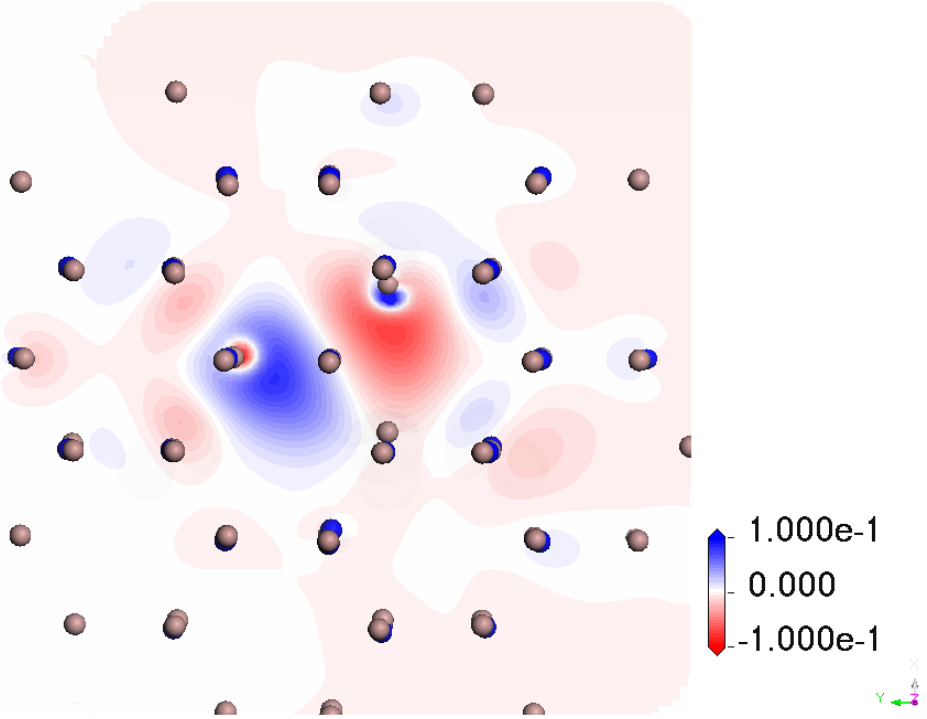


Figure 93: The isocontours of the charge density of the highest occupied orbital of the charge 1- nitrogen vacancy, in e/Å^3, xy-plane.

Table 36: The geometries for the nitrogen vacancy. The initial symmetry reduced by off-centre displacement. The inequivalent neighbour lies along the c-axis.

Chargel	Distance between defect centre and nearest Ga (Å)
1+	1.997 (2 equivalent) 2.129 (inequivalent), 1.853 (1 equivalent)
0	1.944 (2 equivalent) 1.955 (inequivalent), 1.758 (1 equivalent)
1-	1.916 (2 equivalent) 1.863 (inequivalent), 1.673 (1 equivalent)

Finally, the distances of the nearest-neighbour Ga atoms to the defect centre (site of the original anion) are given in Table 36. The symmetry is lowered by the presence of the defect. In the bulk, three of the nearest neighbours are equidistant to a nitrogen, with the fourth, inequivalent, nearest neighbour, along the c-axis, at a slightly different distance. When the nitrogen is removed, two Ga ions remain equivalent, while the third one distorts. This symmetry lowering is not observed in the calculations in [32].
Table 36 also suggests that as the charge state becomes more negative, the neighbouring Ga ions move inwards as can be expected due to their positive charge.
Ref. [33] also observes the symmetry lowering distortions. This is attributed to the bonding between the Ga nearest neighbours of the vacancy and is a manifestation of the Jahn-Teller effect (see 3.8.3).

8.4 Discussion

In this section, the implications of the results from the previous sections will be discussed. First, we consider the sources of error in our calculations and compare them with results using periodic boundary condition. Next, we discuss the relative formation energies of the different types of defect under various conditions and the implications of our defect formation energy results for the p-type vs. n-type nature of GaN. We then turn our attention to summarising the thermodynamic defect levels calculated previously

and compare them with deep level transient spectroscopy (DLTS) data. The optical defect levels are compared with optically detected magnetic resonance (ODMR) and photoluminescence data, yellow luminescence at 2.2eV, in particular. The QM/MM results are then related to our previous MM calculations.

8.4.1 Sources of error

Several aspects contribute to inaccuracies in the calculations of defect properties.

The convergence of the calculations with respect to several model parameters has been discussed previously. Region I (QM region) size is an important factor and is related to the magnitude of the boundary effects. Increasing the QM region size, on the one hand, removes the boundary further away from the defect centre, in theory improving the calculation. However, if there is a sizeable lattice mismatch between the optimised lattice parameters calculated with the MM and QM models, increasing the QM region size will tend to exacerbate the problem by increasing internal strain and lead to significant boundary distortions.

The convergence of bulk IPs with respect to the size of Region I was discussed in Section 8.3.1. The Ga vacancy defect formation energies and defect levels, calculated with two cluster sizes, can be found in Section 8.3.2. From the results we can put a tentative estimate on the errors arising from this source of a few 0.1eV at most.

There are further convergence errors due to the size of region III. As discussed in Section 8.2.3, the errors due to this are estimated to be less than 0.1eV. Finally, the errors due to the convergence of the GAMESS-UK calculation are typically very small, less than 0.01eV.

Another important model parameter is the energy functional. In Section 8.3.2, the hybrid functional used here, B97-2 and is compared with an LDA functional SVWN. The differences in calculating the Ga vacancy defect formation energies and optical defect levels were relatively small , a few 0.1eV.

Hybrid functional are considered better quality than LDA functionals.
The largest inaccuracies we found in our calculations arise from the errors due to the insufficiently extensive basis set for the Ga ions. While this choice allows for increased speed and reduced cost of these very expensive calculations, the errors have been found to be of the order of about 3eV for Ga vacancies and interstitials. Future calculations with improved basis sets, for example, the Peterson TZV basis set [46], will improve the accuracy.

8.4.2 Comparison with PBC results

In this section, we will address the question of why such large differences are observed between our calculations and the PBC results. Significant discrepancies are observed in particular for the negatively charged Ga vacancies, which were predict to be much higher in energy than PBC calculations at the CBM. Furthermore, we predict lower energies for the positively charged Ga interstitials and N vacancies at the VBM. In addition to the basis set issues discussed in the previous section, different choices of boundary conditions and differences in energy functionals are two important candidates for the large discrepancies and are discussed below.
Firstly, the periodic boundary conditions lead to self-interactions of defect images and elastic effects due to incomplete relaxation of the geometry around the defect. The cluster methodology in the present work artificially confines electrons and the "tails" of the wavefunctions might not be represented accurately. More delocalised states generally suffer from the QM/MM boundary conditions more than highly localised ones. Furthermore, distortions at the QM/MM boundary also lead to errors, although assuming the distortions are sufficiently far away from the defect centre they can be treated as perturbations to the Hamiltonian which largely cancel out in the calculation of the relevant energies.
The convergence of the relevant energies with the size of the supercell in the PBC calculations and with the size of the QM cluster in QM/MM methodology can give an estimate of the errors due to these effects. The errors in

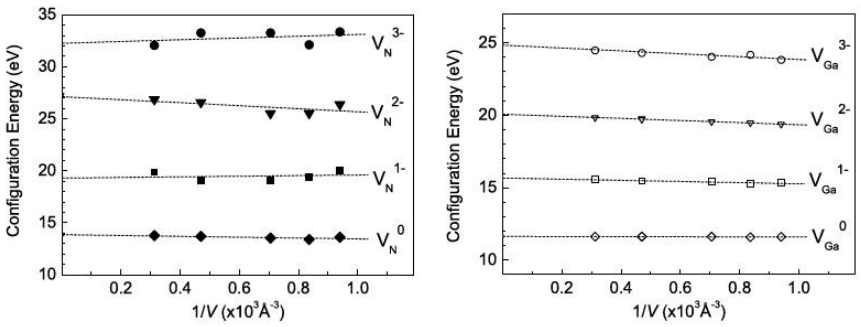


Figure 94: The convergence of the configuration energy of N and Ga vacancies using local spin density approximation (LSDA) and PBCs. From [33].

our calculations were considered in the previous section. For PBC calculations, such a study was performed, for example, in [33]. Their convergence results for Ga and N vacancies are shown in Fig.94. The errors are within about 1eV and more pronounced for the N vacancy. [25] finds the differences in formation energies between PBC calculations with 32-atom and 96-atom supercells to be of the order of a few 0.1eV.

Beyond the problems with the use of boundary conditions, the PBC calculations suffer from problems due to the use of either LDA or GGA energy functionals, which significantly underestimate the band gap: with LDA the band gap is 2.20eV and GGA gives 1.67eV [44]. Such small values of the band gap may lead to excessive hybridisation between the in-gap defect states and the conduction band states, causing potentially significant errors in the calculated energies or even erroneously predicting a state to be a resonance in the conduction band. Hybrid functionals, such as the one used in the present work, tend to give more accurate estimates of the band gap.

In addition to the band-gap problem, LDA is often thought to be unsuitable for defect calculations involving large relaxations and is generally accepted not to predict bond strengths and lengths as accurately as hybrid functionals, due to LDA assuming a greater degree of electron self-interaction, which leads to a more delocalised nature of the electrons. This problem becomes particularly acute when modelling very localised states. GGA is often not found to offer substantial improvement over LDA in this respect and indeed

LDA is sometimes thought superior in predicting properties of GaN [33]. To separate the band-gap problem from the issues with modelling bond strengths with LDA functionals, we performed a QM/MM calculation of the Ga vacancy with 42 QM atoms using the SVWN and B97-2 (as used in the rest of this work) energy functionals. The results, in Section 8.3.2, show small difference in the Ga vacancy formation energies calculated with LDA and B97-2 functionals - a few 0.1eV, which is significantly less than the discrepancies observed between PBC LDA results in [25] and [33], and supports our conjecture that the band gap problem is the main cause of the difference.

In addition to the issues mentioned above, early PBC calculations tend to use non-spin polarised methods, meaning that spin singlets are always calculated for systems with even numbers of electrons and the possibility of triplets or higher multiplets is excluded. For open shell systems, a spin doublet is formed, with the α and β orbitals half-occupied, which might potentially to inaccuracies in the calculation of the formation energy. Later works, such as [33], use the local spin density approximation with PBCs and find differences in formation energies of about 1eV with respect to the results in [25] although other aspects of their set-up differ as well such as the method of treating the 3d electrons and the size of the supercell so it is difficult to establish what effect the spin polarisation had on the result.

In summary, we argue that the choice of less accurate energy functionals based on the local density approximation in the PBC studies of GaN native defects to date leads to errors. In particular, LDA tends to underestimate the band gap seriously and leads to unphysical hybridisation of the in-gap states and the conduction band. We speculate in the next section that this might be the main cause of the substantial difference between the PBC [25] and our calculations of the Ga vacancy at the CBM even considering the improvements to our basis set. In future work, hybrid functionals, such as HSE06, may be applied to the study GaN native defects to improve the results further.

Our own results suffer from inaccuracies as outlined in the previous section, and our formal charge hypothesis may skew results. Further improvements, especially with respect to convergence of the QM region size and refinements

in the basis set, will be implemented in future work. Our formal charge hypothesis could be tested by building a partial charge model and comparing the results.
Ultimately, the best test of the accuracy of the various models is comparison with experimental results, discussed in the following sections.

8.4.3 Defect formation energies and p-type vs. n-type doping

The formation energies of the most stable charge state for all four types of native defect are summarised in Table 37 both at the VBM and CBM.
At the CBM, we predict the N vacancy to be the most stable defect in the Ga-rich limit, in contrast to [25] (PBC DFT calculation with 96-atom supercells under the LDA), which suggests Ga vacancies are the major defect under both Ga- and N-rich conditions, with a formation energy for the 3- charge at the CBM of about 1.5eV in the Ga-rich limit - 7eV lower than our calculation, 3.5eV after improving our basis set as described in the discussion of the Ga vacancies and interstitials. Using our most recent calculations with a more extensive basis set on Ga, as discussed earlier in this chapter, the Ga vacancies appear to be the most stable species in the N-rich limit. The PBC results in [33] smilarly suggest that Ga vacancies are the most stable species in N-rich conditions, with very similar formation energies for both type of vacancy in Ga-rich conditions.
The formation energy of the N vacancy is predicted to be 3.6eV at the CBM in [25] for the 1+ charge state, while our calculations give 2.3eV.
[34] studied convergence using up to 300 atoms in the supercell and predicts the formation energies of both Ga and N vacancies to be around 0.5eV in the Ga-rich limit at the CBM, much lower than our calculations. Additionally, our calculations do not predict the highly negatively charge states (2-, 3-) of the N vacancy to be favourable unlike the results in Ref. [34].
We speculate that the Ga vacancies might be artificially stabilised by hybridisation of the in-gap states with the conduction band due to the band gap problem in the PBC LDA approach in [25] and [34] as discussed in Section 8.4.2.

Table 37: A summary of lowest energy charge states of each type of native defect in GaN. If the lowest energy charge state is a resonance (RES), the lowest energy non-resonant state is also indicated. The N vacancy 3+ is possibly a resonance with a defect level 0.06eV. Given the error of our calculation it is difficult to determine with certainty whether this is a resonance. S=spin singlet, D=doublet, T=triplet, Q=quadruplet. All calculations were carried out with a 74-QM-atom cluster except the Ga vacancy, which used 116 QM atoms. The values in brackets indicate the results with the improved basis set.

Defect type	Charge state	Ga-rich limit	N-rich limit
VBM			
Ga interstitial	3+	-4.27(-1.86)	-3.05(-0.63)
N interstitial	3+ (RES)	1.32	0.10
	2+	1.90	0.67
N vacancy	3+ (possibly RES)	-3.53	-2.31
Ga vacancy	0(Q)	10.33	8.73
CBM			
Ga interstitial	1+	4.48	5.71
N interstitial	0	5.85	4.63
N vacancy	1+(S)	2.30	3.52
Ga vacancy	3-	8.11 (4.62)	6.88 (3.39)

The formation energies at the VBM become negative for the positive charge states of the N vacancy and Ga interstitial. As has been argued in [43] for the case of ZnO, negative formation energies of positively charged native defects (i.e. cation interstitials and anion vacancies) at the VBM can be interpreted as the defects acting as compensating centres, removing holes. This observation can be understood with reference to the reaction equation corresponding to the formation energy of the defect at the VBM.

$$\mathrm{GaN(bulk)} \rightarrow \mathrm{GaN(with\ N_{va}^{3+})} + 1/2\mathrm{N_2(g)} + 3\mathrm{e^-(VBM)} \tag{110}$$

This equation can be rewritten as

$$\mathrm{GaN(bulk)} + 3\mathrm{h^+(VBM)} \rightarrow \mathrm{GaN(with\ N_{va}^{3+})} + 1/2\mathrm{N_2(g)} \tag{111}$$

Table 38: The formation energies per hole or electron for GaN and ZnO, indicating the reaction energies for the charge carrier compensation processes. The values in brackets refer to the calculations using a larger basis set as discussed in the text. For ZnO, the data is taken from [43].

	Formation energy per electron/hole (eV)					
	GaN			ZnO [43]		
Defect type	n	Ga-rich	N-rich	n	Zn-rich	O-rich
hole carriers (VBM)						
Cation interstitial	3+	-1.42 (-0.41)	-1.02 (-0.01)	2+	-2.35	-0.50
Anion vacancy	3+	-1.18	-0.77	2+	-2.59	-0.74
Anion interstitial	2+	0.95	0.34	2+		
electron carrier (CBM)						
Cation vacancy	3-	2.70 (1.54)	2.29 (1.13)	2-	2.60	0.75
Anion interstitial	1-	5.88	4.66	2-	3.58	1.73
Anion vacancy	1-	3.42	4.65	1-	unstable	

illustrating the effect of the N vacancy formation on the removal on holes (h). Since the formation energy is exothermic, holes in the valence band would not be present in appreciable concentrations in thermodynamic equilibrium, which explains the difficulty in p-doping both GaN, and even more so ZnO. The relevant formation energies per hole or electron for the two materials are compared in Table 38. The ZnO values are based on the formation energies of the formal charge defects from [43]. The GaN values are derived from the lowest energy non-resonant charge state, positive charge state for the calculation of compensation processes involving hole carriers, and negative for the equivalent calculation related to electron carriers.

We propose that the exothermic formation energies of cation interstitials and anion vacancies are the cause of the challenges of p-doping both semiconductors. The hole compensation energies in ZnO are up to about 1.4eV more exothermic for cation-rich conditions, due to the high lattice energy of ZnO, whilst under anion-rich conditions GaN formation energies are more exothermic by up to 0.5eV. ZnO has proved to be even more challenging to p-dope than GaN and our findings suggest this should particularly be the

case under cation-rich growth conditions. All electron compensation processes are endothermic, as can be expected for a naturally n-type material. Ref. [25] also predicts exothermic formation energies for the N vacancy at the VBM of about -1eV in the 3+ charge state, about 1.3eV higher than our prediction. Our predictions for the Ga interstitial exothermic formation energies at the VBM are not reproduced in [25], which predicts the 3+ state to have a formation energy of about 2.3eV at this Fermi level.
GaN hole conductivity can be increased significantly by doping with Mg. If the dopant is present in large enough concentrations, it will have a significant effect on the band structure of the doped material, pushing the valence band up due to the changed Coulombic interactions in the case of doping with a dopant such as Mg, which has a less positive charge than Ga. The defect levels in the band gap may be significantly affected by the upward movement of the valence band and hence the thermodynamics of the system may allow p-doping to occur. Epitaxial strain might also have an effect on the energetics of the system. These questions are possible avenues for future research.

8.4.4 Thermodynamic transitions

The thermodynamic transition levels are summarised in Table 39 and Fig. 96. Unlike the results in [25], summarised in Fig. 25 on p. 104, our calculations do not predict negative-U defects (see Section 3.8.3) for the Ga interstitial and N vacancy. Instead of the 3+/1+ thermodynamic transition associated with the negative-U defects in [25], we observe closely spaced 3+/2+ and 2+/1+ transitions in both cases. For the N vacancy, [33] predicts quite different transitions of 1+/1- and 1-/3- at 2.09eV and 2.80eV, respectively.

Another significant discrepancy between our results and the periodic DFT calculations is the position of the thermodynamic transition levels for the Ga vacancy, which are calculated in [33] to be 0.43eV (0/1- transition), 0.77 eV (1-/2-) and 1.07eV (2-/3-) above the VBM. Our calculations place them at 2.31eV, 2.73eV and 3.15eV, respectively. We reiterate the hypothesis that the difference is due to the hybridisation between the in-gap states and the

Table 39: A summary of the in-gap thermodynamic transition levels. Values in eV.

Defect	Transition	Defect level	
		above VBM	below CBM
Ga interstitial	3+/2+	2.56	0.95
	2+/1+	2.68	0.82
N interstitial	3+/2+	0.58	2.93
	2+/1+	1.13	2.38
	1+/0	2.82	0.69
N vacancy	3+/2+	1.12	2.39
	2+/1+	1.21	2.30
Ga vacancy	1-/0	2.31	1.20
	2-/1-	2.73	0.78
	3-/2-	3.15	0.36

conduction band in the PBC calculations. However, further QM/MM calculations with an improved basis set need to be carried out to confirm our values.

The other thermodynamic transition levels are in broad agreement with the PBC values, apart from the 1-/0 and 0/1+ transitions of the N interstitial. The different types of transition and the experimental methods that can access them are shown in Fig. 96 on p. 246 . Among the experimental data that can give insight into defect energetics is deep level transient spectroscopy (DLTS). It has been suggested that the thermodynamic transition level is the relevant quantity measured by this technique [27], [28]. Various defect levels have been reported (values below the conduction band): 0.62eV [30], 0.49eV [31], 0.44eV [29], 0.26eV [30], 0.21eV [29] , 0.20eV [29] and 0.18eV [31]. DLTS studies by Haase et al. [41] have shown that a defect (with activation energy 0.67eV in their experiment) can be generated by N implanatation and then removed by annealing. They speculate that this points to the nitrogen vacancy or interstitial defect.

Within the error of our calculation, there are a number of thermodynamic transition levels in the range of 0.1-1.0eV below the CBM that could explain the experimental observations. Ga interstitial (3+/2+, 2+/1+), N intersti-

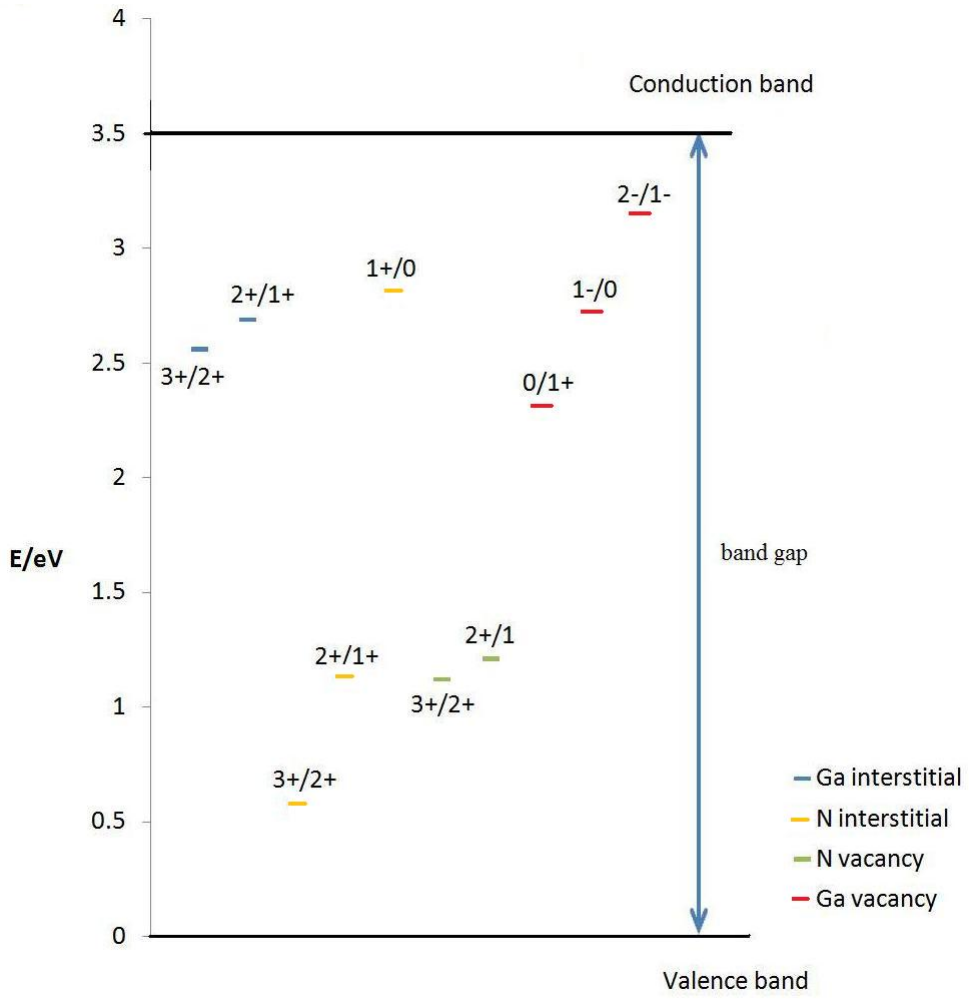


Figure 95: The thermodynamic defect levels in GaN. Values in eV.

tial (1+/0) and Ga vacancy (2-/1-, 3-/2-) are all possible candidates. The N interstitial 1+/0 transition is a promising candidate for interpreting the aforementioned result in [41].

Our preliminary results indicate that much of the correction to the Ga vacancy results due to the increased accuracy of the basis set, as discussed in the previous sections, might come from the improvement in the Ga free atom energy, in which case the formation energies would be offset by a constant amount for the different charge states and the thermodynamic transition levels would be unaffected. A full set of calculations for all charge states is

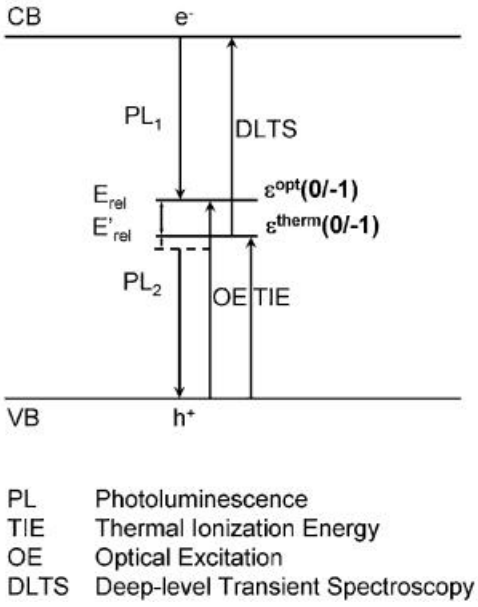


Figure 96: Types of transition level and spectroscopic methods for measuring them, for an example transition, 0/1-. ϵ^{opt} and ϵ^{therm} refer to optical and thermodynamic levels respectively. Upward arrows indicate electron excitation and downward arrows electron decay. The relaxation energies E_{rel} and E'_{rel} and the PL_1 and PL_2 transitions are further described in Fig. 97. From [27]

recommended for Ga interstitials and vacancies to confirm the present values.

8.4.5 Optical transitions

The optical defect levels are summarised in Table 40 and Fig. 98. These levels are potentially useful in interpreting photoluminescence data. One of the most controversial observations is yellow luminescence at 2.2eV.

Fig. 97 shows two types of transitions that may be responsible for luminescence, although of course there are other possibilities, for example transitions between defect states. PL_1 in Fig. 97 describes the electron affinity of the defect with respect to the CBM. PL_2 refers to the ionisation potential with respect to the VBM.

Table 40 suggests the Ga vacancy 1-/0 or 2-/1- PL_2-type transitions as a possible cause for yellow luminescence. Not all electron affinity calculations

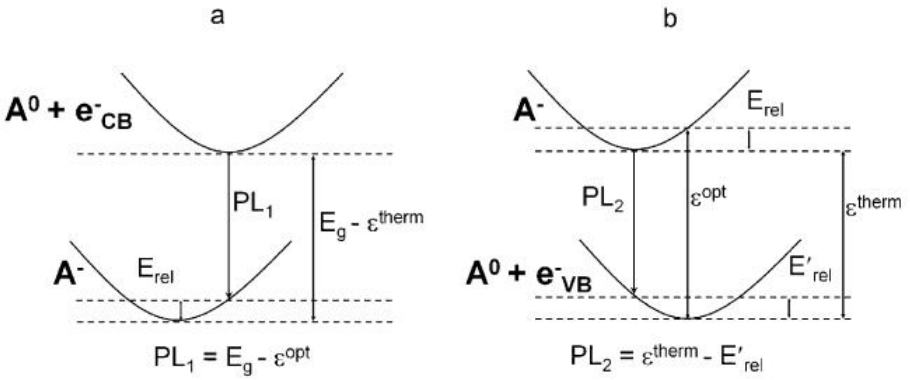


Figure 97: Two types of transitions that can be determined in a photoluminescence measurement. See also Fig.96. From [27]

have been completed but preliminary results suggest that the 2+/1+ affinity of the N vacancy is exothermic with an energy of 1.85eV and this would be another possible candidate although less close to the 2.2eV value than the Ga vacancy results. Further discussion on the controversy regarding the origin of YL can be found in Section 4.3.2. Future calculations to shed more light on this issue are recommended, especially on the issue of defect complexes and C impurities, which have been suggested as a possible cause of YL. The Ga defect results also need to be verified with a larger basis set.
Further luminescence lines and bands in GaN are described in Table 49 on p. 274. With future refinements of our model, more accurate assignments of defect transitions can be made to luminescence lines.
We also note that difference between the optical and thermodynamic defect levels is the relaxation energy associated with the thermodynamic transition.
The optical electron affinities can be used to explain photoluminescence and optically detected magnetic resonance (ODMR) measurements in [26], which indicate a deep donor state about 0.7eV below the conduction band. When this donor is in its initial state it is paramagnetic, i.e. there must be an unpaired spin. When it accepts an electron no ODMR signal arises as the excited state is not paramagnetic. The 0/1- N interstitial transition fits this description as 1- is a spin singlet, 0 is a spin doublet. The electron affinities of the N interstitial are summarised in Table 29. The 0/1- transition is 0.92eV

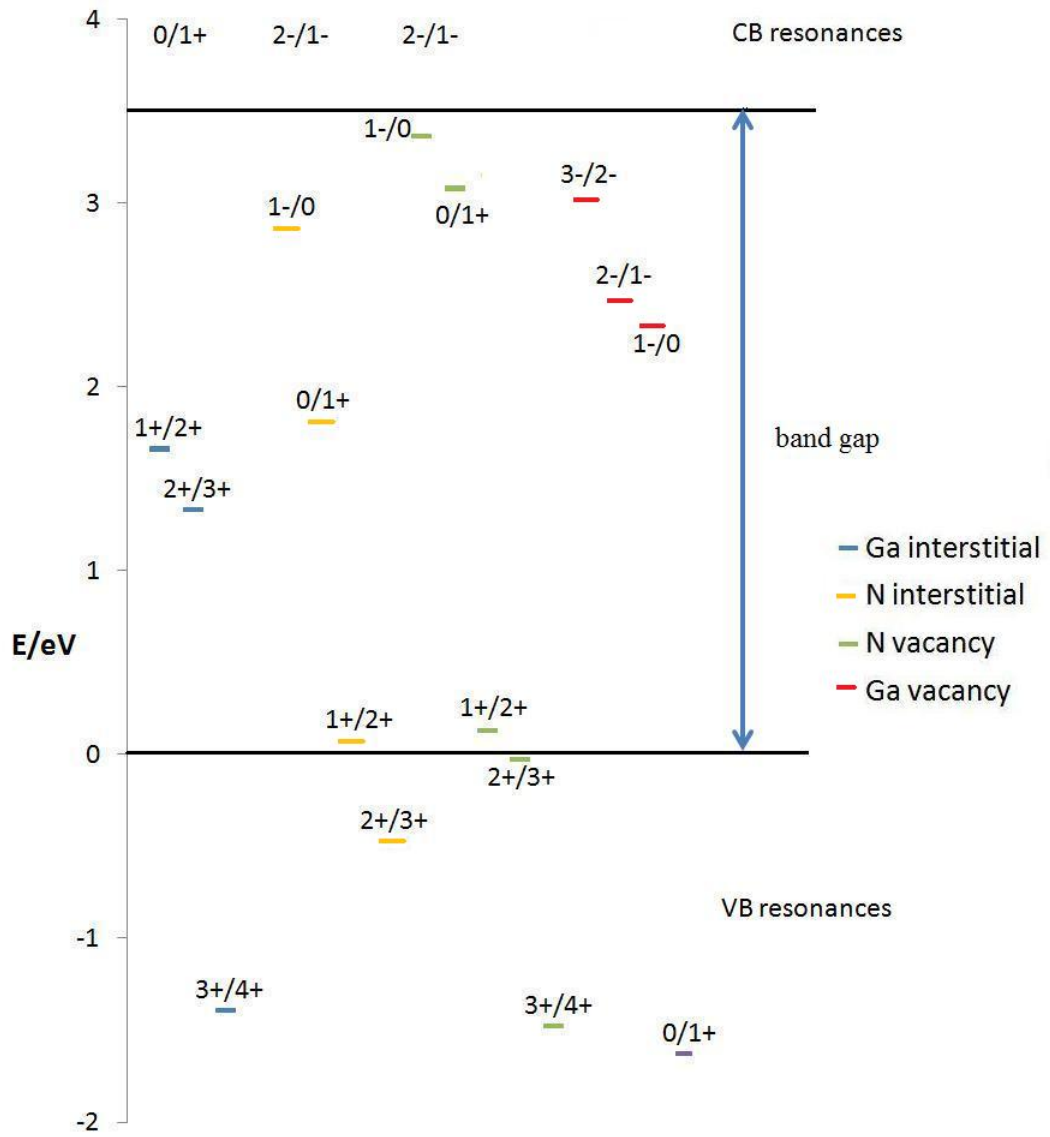


Figure 98: The optical defect levels in GaN.

below the CBM and, within the error of our calculation, a good candidate for the defect which gave rise to the results in [26]. Further calculations on the Ga vacancy optical affinities are to be performed in the future as well as corrections due to the basis set.

Table 40: A summary of the in-gap optical transition levels.

Defect	Transition	Defect level	
		above VBM	below CBM
Ga interstitial	2+/3+	1.30	2.21
	1+/2+	1.63	1.88
N interstitial	1+/2+	0.04	3.47
	0/1+	1.77	1.73
	1-/0	2.83	0.68
N vacancy	2+/1+	0.13	3.38
	0/1+	3.08	0.42
	1-/0	3.33	0.18
Ga vacancy	1-/0	2.33	1.17
	2-/1-	2.48	1.03
	3-/2-	3.03	0.48

8.4.6 Comparison with our MM results

Finally, we briefly compare the results of our MM defect calculations, which were calculated for the formal charge states of the defects. In this chapter, we found the N interstitial in the charge state 3- to be a resonance in the conduction band although further calculations are needed to explore different geometries, particularly a split interstitial configuration with the N in the middle of the hexagonal channel, which was found to be the lowest energy arrangement in our MM defect studies. The N vacancy in the formal charge state is a resonance in the valence band hence again MM calculations are of limited use.

Our QM/MM calculations indicate that the Ga interstitial and vacancy are stable in their formal charge states. The Ga Frenkel defect energy is calculated as 14.36eV/per defect pair, which changes to 13.89 eV/defect pair when the more accurate basis set is used. We attribute the discrepancy between the QM/MM values and the MM result, 19.44eV/defect pair, to the inaccuracies of the MM potential. The distances between defect centres and the nearest neighbours differ by up to about 0.2Å between the MM and QM/MM sets of results, and even more for further neighbours, as can be seen from Tables 9 (p. 150), 24 (p. 214) and 27 (p. 217). The Coulomb potential contributions

will therefore be appreciably different (see Fig. 35 on p. 128).
The inaccuracy of the defect geometries may be due to the parameters of the potential being fitted using bulk properties. The interatomic distances of the defect configurations differ significantly from the bulk distances and the results are therefore less reliable than bulk property calculations. Fitting to a wider range of properties which explore non-equilibrium interatomic distances, or to *ab initio* potential surfaces, might improve the MM results. Finally, further refinements to the QM/MM values due to the use of larger clusters are also possible in the future and might alter the extent of the discrepancy. While MM calculations can be a starting point in the study of GaN defects, they are of limited usefulness in exploring the full range of the complex defect properties of the material.

8.5 Conclusions and future work

In this chapter, we presented a QM/MM model for native defects in GaN. We have shown that our results differ substantially from previous period boundary calculations, especially in predicting significantly higher formation energies for the formally charged Ga vacancy. The variations in the formation energies calculated using periodic boundary conditions and our method are of the order of a few eV. We predict the N vacancy to be the thermodynamically stable species at the CBM in the Ga-rich limit and the Ga vacancy is the likely candidate for the most favourable defect in the N-rich limit. We also do not observe negative-U defects in the case of Ga interstitials and N vacancies, unlike the PBC calculations in [25] and [34]. We attribute these differences to the choice of functional and the band gap problems in the PBC calculations as well as inaccuracies due to the basis set and convergence with respect to the QM region size in our own calculations. Our formal charge hypothesis may also skew results and is a candidate for further investigation in the future.
We further discussed the exothermic formation energies of the Ga interstitial and N vacancy defects as the cause of the difficulties in p-doping GaN.
We also shed some light on the elusive issue of yellow luminescence. Our

results support previous suggestions that Ga vacancies may be the cause of YL in GaN, although N vacancy is another likely candidate. Further calculations on the energetics of extrinsic defects are recommended to establish what role they play in luminescence. In future work, our calculations can be used to assign other GaN bands and lines.

We suggest that the processes responsible for a variety of DLTS signals between 0.18-0.67eV below the CBM may be due to Ga interstitial 3+/2+, 2+/1+ transitions, N interstitials (1+/0) and Ga vacancies (2-/1-, 3-/2-). We attribute the ODMR signal indicating a deep donor state 0.7eV below the conduction band to the N interstitial 0/1- transition.

Experimental results can be a useful test of the accuracy of the theoretical models. However, a number of defect transitions, involving native and extrinsic defects, can be reponsible for any one experimental observation and firm conclusions about the experimental energetics of specific transitions, that can be directly compared to calculated values, are difficult to draw. Very targeted experiments that give information about transitions between specific defect states would be valuable in testing our hypotheses. Alternatively, consistency with a range of experimental data would give credence to our model.

We have identified improvements to our model for Ga vacancies and interstitials for future calculations. Potential further work also includes exploring different configurations of the native defects, especially the channel-centred configuration of the N interstitial. Extrinsic defects, such as C, O and Mg, and defect complexes may play an important role in understanding the properties of this material and provide exciting opportunities for future study.

References

[1] Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials, A.A.Sokol, S.T.Bromley, S.A.French, C.R.A.Catlow, P.Sherwood, Int.J.of Quantum Chem. 99 **695 - 712** (2004)

[2] Hartree-Fock cluster computations of defect and perfect ionic crystal properties, J. H. Harding, A. H. Harker, P. B. Keegstra, R. Pandey, J. M. Vail and C. Woodward, Physica B+C 131 **151 - 156** (1985)

[3] QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis, P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjvoll, A. Fahmi, A. Schfer, Ch. Lennartz, J. Mol. Struct. (Theochem.) 632 **1 - 28** (2003)

[4] Identication and characterization of active sites and their catalytic processesthe Cu/ZnO methanol catalyst, S. A. French, A. A. Sokol, S. T. Bromley, C. R. A. Catlow, P. Sherwood, Topics in Catalysis 24 **161 - 172** (2003)

[5] Relative energies of surface and defect states: ab initio calculations for the MgO (001) surface, P.V.Sushko, A.L.Shluger, C.R.A.Catlow, Surface Sci. 450 **153 - 170** (2000)

[6] `http://www.cfs.dl.ac.uk/docs/index.shtml`

[7] The GAMESS-UK electronic structure package: algorithms, developments and applications, M.F.Guest, I.J.Bush, H.J.J.Van Dam,P.Sherwood, J.M.H.Thomas, J.H.Van Lenthe, R.W.A.Havenith, J.Kendrick, Molecular Physics 103 **719 - 747** (2005)

[8] `http://www.hector.ac.uk/`

[9] Nitride semiconductors and devices, H.Morkoc, Springer (1999)

[10] V. Bougrov, M.E.Levinshtein, S.L.Rumyantsev , A.Zubrilov , in Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe. Eds. Levinshtein M.E., Rumyantsev S.L., Shur M.S., John Wiley & Sons, Inc., New York, (2001) **1 - 30**

[11] CRC Handbook of Chemistry and Physics, 73rd Ed., edited by David R. Lide,CRC Press, Boca Raton, (1992) **518**

[12] Handbook of Chemistry and Physics, D.R.Lide, 89th Edition, 2008-2009

[13] Hybrid exchange-correlation functional determined from thermochemical data and ab initio potentials, P.J. Wilson, T,J. Bradley, D. J. Tozer, J. Chem. Phys. 115 **9233 - 9242** (2001)

[14] Relativistic compact effective potentials and efficient, shared exponent basis sets for the third-, fourth- and fifth- row atoms, W.J. Stevens, M. Krauss, H. Basch, P. G. Jasien, Can. J. Chem. 70 **612 - 630** (1992)

[15] Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy 7, F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys., 3297 **3297 - 3305** (2005)

[16] Dr. Alexei Sokol, private communication

[17] Electronic structure of ZnO:GaN compounds: Asymmetric bandgap engineering, M. N. Huda, Y. Yan, S.-H. Wei, M.M. Al-Jassim, PRB 78 **195204** (2008)

[18] Band alignment at a ZnO/GaN (0001) heterointerface, S.-K. Hong,T. Hanada, H. Makino, Y.Chen, H.-J. Ko,T. Yao, APL 78 **3349 - 3351** (2001)

[19] Band offsets of polar and nonpolar GaN/ZnO heterostructures determined by synchrotron radiation photoemission spectroscopy basic solid state physics, J. W. Liu, A. Kobayashi, S. Toyoda, H. Kamada, A. Kikuchi, J. Ohta, H. Fujioka, H. Kumigashira. M. Oshima, Phys. Status Solidi B, 1-4 **956 - 959** (2010)

[20] Effects of counterpoise correction and basis set extrapolation on the MP2 geometries of hydrogen bonded dimers of ammonia, water, and

hydrogen fluoride, A. D. Boese, G. Jansen, M. Torheyden, S. Höfener, W. Klopper, Phys. Chem. Chem. Phys.13 **1230 - 1238** (2011)

[21] Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, S. H. Vosko, L. Wilk, M. Nusair, Can. J. Phys. 58 **1200 - 1211** (1980)

[22] Interaction of hydrogen with nitrogen interstitials in wurtzite GaN, A.F.Wright, J.Appl.Phys. **6526 - 6532** (2001)

[23] `http://cnx.org/content/m14777/latest/`

[24] `http://en.wikipedia.org/wiki/File:MO_diagram_pi_orbitals.png`

[25] Diffusivity of native defects in GaN, S.Limpijumnong, C.G.van de Walle, Phys.Rev.B 69 **035207** (2004)

[26] Optically detected magnetic resonance of GaN films grown by organometallic chemical vapour deposition, E.R.Glaser, T.A.Kennedy, K.Doverspike, L.B.Rowland, D.K.Gaskill, J.A.Freitas Jr., M.Asif Khan, D.T.Olson, J.N.Kuznia, D.K.Wickenden, Phys.Rev.B 51 **13326 - 13336** (1995)

[27] *Ab initio* study of transition levels for intrinsic defects in silicon nitride, C.Di Valentin, G.Palma,G.Pacchioni, J.Phys.Chem.C 115 **561 - 569** (2011)

[28] First-principles calculations for defects and impurities: Applications to III-nitrides, C. G. Van de Walle J. Neugebauer, J.Appl.Phys. 95 **3851 - 3879** (2004)

[29] Deep centers in n-GaN grown by reactive molecular beam epitaxy, Z-Q. Fang, D. C. Look, W. Kim, Z. Fan, A. Botchkarev, H. Morkoc, APL 72 **2277 - 2279** (1998)

[30] Deep levels in the upper band-gap region of lightly Mg-doped GaN, P. Hacke, H. Nakayama, T. Detchprohm, K. Hiramatsu, N. Sawaki, APL 68 **1362 - 1364** (1996)

[31] Deep level defects in n-type GaN, W. Götz, N. M. Johnson, H. Amano, I. Akasaki, Appl. Phys. Lett. 65 **463 - 465** (1994)

[32] Atomic geometry and electronic structure of native defects in GaN, J.Neugebauer, C.G. van de Walle, Phys.Rev.B 50 **8067 - 8070** (1994)

[33] Vacancies in wurtzite GaN and AlN, K. Laaksonen, M. G. Ganchenkova, R. M. Nieminen, J. Phys. Condens. Matter 21 **015803** (2009)

[34] Nitrogen vacancies as major point defects in gallium nitride, M.G.Ganckenkova, R.M.Nieminen, PRL 96 **196402** (2006)

[35] Ionisation energy of GaN point defects, A.Bere, I.Belabbas, G.Nouet, P.Ruterana, J.Chen. G.Segda, J.Koulidiati, phys.stat.sol.(c) 4 **2593 - 2596** (2007)

[36] Intrinsic defect properties in GaN calculated by *ab initio* and empirical potential methods, F.Gao, E.J. Bylaska, W. J. Weber, Phys.Rev.B **245208** (2004)

[37] Calculated defect levels in GaN and AlN and their pressure coefficients, I. Gorczyca, A. Svane, N. E. Christensen, Solid State Communications 101 **747 - 752** (1997)

[38] Native defects in Gallium Nitride, R.Boguslawski, E.L.Briggs, J.Bernholc, Phys.Rev.B 51 **17255 - 17258** (1995)

[39] Point-defect complexes and broadband luminescence in GaN and AlN, T. Mattila, R. M. Nieminen, Phys. Rev. B 55 **9571 - 9576** (1997)

[40] Nonlinear ionic pseudopotentials in spin-density-functional calculations, S. G. Louie, S. Froyen, and M. L. CohenPhys. Rev. B 26 **1738 - 1742** (1982)

[41] Deep-level defects and n-type-carrier concentration in nitrogen implanted GaN, D. Haase, M. Schmid, W. K ürner, A. Dörnen, V. Härle, F. Scholz, M. Burkard, H. Schweizer, APL 69 **2525 - 2527** (1996)

[42] Simulation of GaN using an environment-dependent empirical tight-binding model, D.E.Boucher, G.G.DeLeo, W.B.Fowler, Phys.Rev.B 59 **10064 - 10070** (1999)

[43] Microscopic origins of electron and hole stability in ZnO, C.R.A.Catlow, A.A.Sokol, A.Walsh, Chem.Comm. (in press)

[44] Origins of Fermi-level pinning on GaN and InN polar and nonpolar surfaces, D. Segev, C. G. Van de Walle, Europhys. Lett. 76 **305 - 311** (2006)

[45] Gallium vacancies and the yellow luminescence in GaN, J.Neugebauer, C.G. Van de Walle, APL 69 **503 - 505** (1996)

[46] Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 1618 elements, K.A. Peterson, J. Chem. Phys. 119 **11113 - 11123** (2003)

[47] Small-core multiconfiguration Dirac-Hartree-Fock-adjusted pseudopotentials for post-d main group elements: Application to PbH and PbO B. Metz, H. Stoll, M. Dolg, J. Chem. Phys. 113 **2563 - 2569** (2000).

[48] `https://bse.pnl.gov/bse/portal`

9 Conclusions

This work has presented two models for GaN: a molecular mechanics (MM) model and a hybrid QM/MM (quantum mechanics-molecular mechanics) one.

The MM parameterisation is a formal charge model for simulating the properties of GaN, as well as the related materials AlN and InN, using multilayered Buckingham and Morse potentials and the shell model for describing the polarisability of the nitride ion. The AlN and InN N-N interaction parameters are fixed at the values developed for GaN, apart from the spring constant of InN, which had to altered to allow the dielectric constants to be modelled correctly. The common N-N interaction allowed us to study the properties of alloy systems. Overall, the structural and dielectric properties are well reproduced.

The main shortcoming of the MM model is the overestimation of the off-diagonal elastic constants, meaning the model has a lesser tendency to shear than the physical material. This deficiency affects physical properties such as compressibility along the c-axis. Because the phase transition between the wurtzite and rocksalt phases is a first-order transition, it is less significantly impacted, and it can be fitted well within our model by shifting the separate layers of our potential by constant offsets.

The origin of the problems with modelling the off-diagonal terms, which have persisted despite intensive efforts at improving the model, is probably the significant deviation of GaN from a fully ionic material. The formal-charge assumption is based on physical reasoning [1] and the practical considerations for implementing the QM/MM scheme. An alternative to introducing covalency into the scheme via a partial-charge model, which would require significant changes to QM region termination, is to include three-body terms and future work on incorporating these terms could improve the model of the shear response. We did not attempt this in the present work as the main focus was the development of a QM/MM model, where the shear response is of little consequence.

The model was then used to calculate the properties of formal charge defects,

predicting vacancies to be the predominant type of defect in AlN, GaN and InN.

As discussed above, the common N-N set of interactions allows us to model alloys of the III-V nitrides. Two methods were used, the mean field approach and explicit energy calculations of specific examples of cation ordering.

The mean field approach is most applicable to situations where the alternative cation species are similar in their ionic radii, such as in the case of Al and Ga. The ionic radius of In is much larger than the other two cations and this method is more likely to cause difficulties. It has indeed been found that modelling the structural parameters of InGaN and AlInN with this approach produced less consistent results than the AlGaN model.

Of the InGaN explicit cation ordering arangements we considered, we found that layered configurations, with alternating planes of unlike cations in the *c*-direction, had higher formation enthalpies relative to other structures we studied. In the modelling of alloys, the inaccuracies in the off-diagonal elastic constants might play a role. However, we argue that the layered arrangements are less likely to relieve strain by shearing and therefore not likely to be brought into thermodynamic stability. We find staggered arrangements of unlike cations to be the most thermodynamically stable out of all the configurations we considered, in agreement with previous DFT calculations. In AlGaN alloys, the alternating layer arrangement was found to be more energetically stable.

As was the case with the mean field approach, the enthalpy of mixing of the bulk alloys was found to be positive. Furthermore, we found that the greater the ionic radius mismatch between the cations, the larger the calculated enthalpy of mixing, as could be expected.

There has been a suggestion, based on experimental x-ray diffraction studies, that *c*-axis ordering is present in alloys of AlGaN and InGaN. We suggest that these might not necessarily be the thermodynamically stable configuration but rather a consequence of the growth process. Indeed, the processes at the growing surface warrant further study.

Epitaxial strain was found to have a stabilising effect on the structures, particularly in the case of the more internally strained alloys, such as InGaN, in

agreement with experiment. AlGaN showed little stabilisation with epitaxial strain in our calculations.

We also found a dependence of several quantities such as the enthalpy of mixing and structural parameters on the number of unlike nearest neighbours in an alloy. To the best of our knowledge, this has not been reported before.

Future improvements of the alloy model could include improving the modelling of the spring constant. The InN N spring constant has a different value at present from the AlN and GaN values, which results in difficulties when modelling the alloy. In the present work, we used a simple averaged model for the alloy N spring constant but have suggested possible improvements (see section 7.3.1 on p. 7.3.1) which can be used to refine the model further.

In summary, we developed and validated a new formal-charge interatomic potential for GaN, AlN and InN, which successfully predicts a number of physical properties of binary nitrides and their ternary alloys. With this model, and its limitations in mind, more complex phenomena can be investigated in the future.

The primary aim of this project was to implement a QM/MM scheme for the GaN system and use it to study native defects in the material. The hybrid method proved more challenging to implement than some oxide systems previously studied (such as ZnO). The main difficulty we encountered was the distortion of the QM/MM boundary due to the mismatch between the QM and MM forces. The process of refitting the MM potential to remove these difficulties is very time-consuming and we suggest an automated method (i.e. a fitting code) is developed if the methodology is to be used more widely for complicated solid state systems in the future.

Our QM/MM defect calculations differ in several respects from previous PBC DFT results, especially in predicting high formation energies for the Ga vacancy at the CBM. Our calculations show N vacancies to be the dominant defect at the CBM, at least in the Ga-rich limit. We also propose Ga or N vacancies to be responsible for the commonly observed, and controversial, yellow luminescence in GaN. Furthermore, the exothermic formation energies for the Ga interstitial and N vacancy at the VBM are used to explain the problems with p-doping the material. We also make predictions about the

defect transitions responsible for the observations in the recent DLTS and ODMR experiments.

Further investigations of the differences between QM/MM and PBC results are necessary to understand the causes of the discrepancy between these two sets of results. We suggest that the differences might be due to the rather inaccurate LDA functional used in the PBC GaN calculations to date, which tends to underestimate the band gap quite severely. A recent study on the C impurity in GaN [2] with a PBC approach and the HSE06 functional could provide a basis for comparison. A PBC study of native defects with this functional would shed light on the issue.

We have also suggested improvements to our Ga basis set to be included in future calculations; these might also explain some of the differences between our results and PBC calculations. The choice of basis set has to balance carefully the need for a sufficiently large number of basis functions to model each QM ion with the computational expense of the calculations, which rises rapidly with the size of the QM region, another important determinant of the quality of the results. Convergence of the defect energies with QM region size should also be investigated further. We recommend the formal charge hypothesis is revisited and a comparison with a partial charge QM/MM scheme is carried out for completeness.

With the methodological advances put forward in this thesis, many applications can be explored, including surfaces, interfaces and defect migration. Extrinsic defects in GaN also offer many exciting opportunities, especially Mg, which has been used to create p-doped GaN. The present model could be used to understand the mechanism of hole creation. The AlN and InN potentials developed here are candidates for new QM/MM studies.

A relatively recent development in QM/MM modelling is the so-called Learn-On-The-Fly (LOTF) [3], which aims to reduce the boundary effects in QM/MM calculations. This method does not conserve energy and further developments will be needed to allow the extraction of energetic information. This method is another avenue for future study.

In summary, this work has provided new insights into the physical properties of GaN and related materials, which have many useful industrial applica-

tions, and also implemented and validated the QM/MM methodology for future studies.

References

[1] Ionicity in solids, C.R.A.Catlow, A.M.Stoneham, J.Phys.C:Solid state phys. 16 **4321 - 4338** (1983)

[2] Carbon impurities and the yellow luminescence in GaN, J. L. Lyons, A. Janotti, C. G. Van de Walle, APL 97 **152108** (2010)

[3] "Learn on the Fly" A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation, G. Csanyi, T. Albaret, M. C. Payne, A. De Vita, PRL 93 **175503** (2004)

10 Appendix I - Interatomic potentials

Table 41: The units of the potential parameters in the current potential models.

Interaction	Parameter	Unit	Interaction	Parameter	Unit
Lennard-Jones	C_n	eVÅ^n	Morse	D_e	eV
Buckingham/	A	eV		a	Å^{-1}
Born-Meyer	ρ	Å		r_0	Å
	C	eVÅ^6	polynomial	c_n	eVÅ^{-n}
spring	k_n	eVÅ^{-n}			

Table 42: Interatomic distances for GaN calculated with experimental structural parameters. Rocksalt values calculated by optimising at zero pressure with our potential.

Atom pair	Dist. (Å)	Number of neighbours	Dist. (Å)	Number of neighbours
Wurtzite				
Ga - Ga	3.1787	6	3.1846	6
Ga - N	1.9473	1	1.9487	3
	3.238	1	3.7328	6
	3.7335	3		
N-N	3.1787	6	3.1846	6
zincblende				
Ga - Ga	3.1750	12		
Ga - N	1.9443	4	3.7230	12
N - N	3.1750	12		
rocksalt				
Ga - N	2.1098	6	3.6542	8
N - N	2.9837	12		
Ga - Ga	2.9837	12		

Table 43: Interatomic distances for InN calculated with experimental structural parameters. The rocksalt values were obtained by optimising with our potential.

Atom pair	Dist. (Å)	Number of neighbours	Dist. (Å)	Number of neighbours
Wurtzite				
In - In	3.5030	6	3.5400	6
In - N	2.1527	3	2.1689	1
	3.5211	1	4.1432	3
	4.1516	6		
N-N	3.5030	6	3.5400	6
zincblende				
In - In	3.5214	12		
In - N	2.1564	4	4.1292	12
N - N	3.5214	12		
rocksalt				
In - N	2.3530	6	4.0755	8
N - N	3.3276	12		
In - In	3.3276	12		

Table 44: Interatomic distances for AlN calculated with experimental structural parameters. The rocksalt values were obtained by optimising withour potential.

Atom pair	Dist. (Å)	Number of neighbours	Dist. (Å)	Number of neighbours
Wurtzite				
Al - Al	3.0707	6	3.1100	6
Al - N	1.8886	3	1.9054	1
	3.0766	1	3.6386	3
	3.6473	6		
N-N	3.0707	6	3.1100	6
zincblende				
Al - Al	3.0971	12		
Al - N	1.8966	4	3.6317	12
N - N	3.0971	12		
rocksalt				
Al - N	2.0533	6	3.5554	8
N - N	2.9038	12		
Al - Al	2.9038	12		

Table 45: The parameters of GaN potential. Units in Table 41. Charges Ga +3.0 e, N core 1.4e, N shell -4.4e, For details of functional forms, refer to Section 2.1).

Species	Potential type	Parameters	Range (Å)
Ga core N shell	Lennard Jones 12	$C_{12} = 10$	$0.0 - 15.0$
	Born-Meyer	$A = 1288.5621$	$0.0 - 2.3$
		$\rho = 0.357118$	
	polynomial	$c_0 = 0.55$	$0.0 - 2.3$
	polynomial	$c_0 = 6838.756066$	$2.3 - 2.8$
		$c_1 = -13396.18758$	
		$c_2 = 10494.93436$	
		$c_3 = -4105.622176$	
		$c_4 = 801.4915743$	
		$c_5 = -62.44285935$	
	Born-Meyer	$A = 1043.4481$	$2.8 - 3.4$
		$\rho = 0.384153$	
	polynomial	$c_0 = -0.1$	$2.8 - 3.4$
	polynomial	$c_0 = 134573.1245$	$3.4 - 3.6$
		$c_1 = -193196.8281$	
		$c_2 = 110917.6787$	
		$c_3 = -31832.1139$	
		$c_4 = 4566.565012$	
		$c_5 = -261.9747598$	
N shell N shell	Morse	$D_e = 0.93078550$	$0.0 - 3.7$
		$a = 2.3885$	
		$r_e = 2.44469$	
	polynomial	$c_0 = -4134.971889$	$3.4 - 4.2$
		$c_1 = 5268.685067$	
		$c_2 = -2683.064126$	
		$c_3 = 682.5077857$	
		$c_4 = -86.71401114$	
		$c_5 = 4.401867951$	
	$C_6 r^{-6}$ term	$C_6 = 97.0$	$4.2 - 15.0$
N shell	spring k_2	49.500407	$0.0 - 0.6$
N shell	spring k_4	50 000	$0.0 - 0.6$

Table 46: The parameters of the slightly adapted GaN potential used in the QM/MM calculations. Units in Table 41. Charges Ga +3.0 e, N core 1.4e, N shell -4.4e, For details of functional forms, refer to Section 2.1).

Species	Potential type	Parameters	Range (Å)
Ga core N shell	Lennard Jones 12	$C_{12} = 10$	$0.0 - 15.0$
	Born-Meyer	$A = 1288.5621$	$0.0 - 2.5$
		$\rho = 0.357118$	
	polynomial	$c_0 = 0.3$	$0.0 - 2.3$
	polynomial	$c_0 = 39877.01195$	$2.5 - 2.8$
		$c_1 = -75263.81056$	
		$c_2 = 56781.39881$	
		$c_3 = -21400.8179$	
		$c_4 = 4029.260072$	
		$c_5 = -303.1546416$	
	Born-Meyer	$A = 1043.4481$	$2.8 - 3.4$
		$\rho = 0.384153$	
	polynomial	$c_0 = -0.1$	$2.8 - 3.4$
	polynomial	$c_0 = 134571.7309$	$3.4 - 3.6$
		$c_1 = -193194.8336$	
		$c_2 = 110916.5371$	
		$c_3 = -31831.78729$	
		$c_4 = 4566.518302$	
		$c_5 = -261.9720884$	
	Born-Meyer	$A = 1.0675506$	$3.6 - 10.0$
		$\rho = 0.221570$	
N shell N shell	Morse	$D_e = 0.93078550$	$0.0 - 3.7$
		$a = 2.3885$	
		$r_e = 2.44469$	
	polynomial	$c_0 = -4134.971889$	$3.4 - 4.2$
		$c_1 = 5268.685067$	
		$c_2 = -2683.064126$	
		$c_3 = 682.5077857$	
		$c_4 = -86.71401114$	
		$c_5 = 4.401867951$	
	$C_6 r^{-6}$ term	$C_6 = 97.0$	$4.2 - 15.0$
N shell	spring k_2	49.500407	$0.0 - 0.6$
N shell	spring k_4	50 000	$0.0 - 0.6$

Table 47: The parameters of InN potential I. Units in Table 41. Charges In core +3.0 e ,N core 1.4e, N shell -4.4e. For details of functional forms, refer to Section 2.1).

Species	Potential type	Parameters	Range (Å)
In core In core	LJ 12-term	C_{12}= 20	0.0 − 10
In core N shell	LJ 12-term	C_{12}= 20	0.0 − 10
	Buckingham	$A = 1279.0808$	0.0 − 2.5
		$\rho = 0.392572$	
		$C = 40.323008$	
	polynomial	$c_0 = 0.17$	0.0 − 2.5
	polynomial	$c_0 = 8888.748798$	2.5 − 2.8
		$c_1 = -16504.78407$	
		$c_2 = 12272.57582$	
		$c_3 = -4563.942524$	
		$c_4 = 848.3962086$	
		$c_5 = -63.0483336$	
	Buckingham	$A = 1528.5920$	2.8 − 3.7
		$\rho = 0.388669$	
		$C = 20.844536$	
	polynomial	$c_0 = -0.08$	2.8 − 3.7
	polynomial	$c_0 = -55975.48438$	3.7 − 3.9
		$c_1 = 72777.54561$	
		$c_2 = -37832.13986$	
		$c_3 = 9828.905042$	
		$c_4 = -1276.252061$	
		$c_5 = 66.25990647$	
N shell N shell	Morse	$D_e = 0.93078550$	0.0 − 3.7
		$a = 2.3885$	
		$r_e = 2.44469$	
	polynomial	$c_0 = -4134.971889$	3.4 − 4.2
		$c_1 = 5268.685067$	
		$c_2 = -2683.064126$	
		$c_3 = 682.5077857$	
		$c_4 = -86.71401114$	
		$c_5 = 4.401867951$	
	$C_6 r^{-6}$ term	$C_6 = 97.0$	4.2 − 15.0
N shell	spring k_2	22.8	0.0 − 0.5
N shell	spring k_4	175 000	0.0 − 10.0

Table 48: The parameters of AlN potential I. Units in Table 41. Charges Al core +3.0e, N core +1.4e, N shell -4.4e. For details of functional forms, refer to Section 2.1).

Species	Potential type	Parameters	Range (Å)
Al core N shell	LJ 12-term	C_{12}= 10.421450	$0.0 - 10$
	Born-Meyer	$A = 1644.6091$	$0.0 - 2.2$
		$\rho = 0.333663$	
	polynomial	$c_0 = 0.65$	$0 - 2.2$
	polynomial	$c_0 = 4687.615495$	$2.2 - 2.8$
		$c_1 = -9390.022249$	
		$c_2 = 7524.973473$	
		$c_3 = -3010.247517$	
		$c_4 = 600.5008231$	
		$c_5 = -47.76268013$	
	Born-Meyer	$A = 675.36504$	$2.80 - 3.30$
		$\rho = 0.385367$	
	polynomial	$c_0 = -0.07$	$2.8 - 3.3$
	polynomial	$c_0 = 239338.3051$	$3.3 - 3.5$
		$c_1 = -352957.4505$	
		$c_2 = 208150.2682$	
		$c_3 = -61359.74221$	
		$c_4 = 9041.488172$	
		$c_5 = -532.7635118$	
N shell N shell	Morse	$D_e = 0.93078550$	$0.0 - 3.7$
		$a = 2.3885$	
		$r_e = 2.44469$	
	polynomial	$c_0 = -4134.971889$	$3.4 - 4.2$
		$c_1 = 5268.685067$	
		$c_2 = -2683.064126$	
		$c_3 = 682.5077857$	
		$c_4 = -86.71401114$	
		$c_5 = 4.401867951$	
	$C_6 r^{-6}$ term	$C_6 = 97.0$	$4.2 - 15.0$
N shell	spring k_2	49.500407	$0.0 - 0.5$
N shell	spring k_4	50 000	$0.0 - 10.0$

11 Appendix II - basis set and pseudopotentials for ChemShell calculations

11.1 Gallium

11.2 Region 1 basis set and associated ECP

11.2.1 Basis set

Based on SBKJC VDZ ECP [14] (most diffuse function removed, another function uncontracted relative to the originally published basis set):

L Ga
-0.0017110 113.9000000 -0.0080460
-0.8230360 9.1550000 -0.3574320
0.4586180 6.6330000 0.6637940
1.1618170 2.2780000 0.7136190
L Ga
1.0 2.1230000 -0.0962610
L Ga
1.0 0.1939000 1.0175730
L Ga
1.0000000 0.8818000 1.0000000
D Ga
0.0288770 70.4300000
0.1662530 21.0500000
0.4277760 7.4010000
0.5704100 2.7520000
D Ga
1.0000000 1.0260000
D Ga
1.0000000 0.3907000

11.2.2 ECP

Based on SBKJC [14]
2 10
1
1 -3.8736300 26.7430200
3
0 4.1247200 3.4653000
2 260.7326300 9.1113000
2 -223.9600300 7.8932900
2
0 4.2003300 79.9935300
2 127.9913900 17.3911400

11.3 Region 2 ECP

Based on Stuttgart RLC ECP (no f function), [48]
3 0
1
2 0.000000000 1.000000000
1
2 43.853972000 5.215960000
1
2 156.103390000 4.308904000
1
2 1.031647000 0.496357000

11.4 Peterson TZV basis set

Based on [46]
S Ga
0.000064 11929.2

0.000473 1783.70
0.002055 378.467
0.120290 29.6623
-0.463392 14.2332
0.744384 2.68739
0.450596 1.15096
0.020785 0.256164
-0.009249 0.122874
0.002291 0.053468
S 10 1.00
-0.000018 11929.2
-0.000123 1783.70
-0.000610 378.467
-0.031264 29.6623
0.126869 14.2332
-0.257857 2.68739
-0.286726 1.15096
0.356828 0.256164
0.571150 0.122874
0.245237 0.053468
S Ga
1 0.256164
S Ga
0.122874 1 0.122874
S Ga
1 0.053468
P Ga
0.000117 648.311
0.002102 103.324
-0.081044 21.0066
0.371433 5.03412
0.501425 2.29458
0.240806 1.01078

0.018440 0.326569
-0.002056 0.117329
0.000711 0.041273
P Ga
-0.000025 648.311
-0.000313 103.324
0.014067 21.0066
-0.072662 5.03412
-0.103072 2.29458
-0.049176 1.01078
0.225329 0.326569
0.551737 0.117329
0.379164 0.041273
P Ga
1 0.3095
P Ga
1 0.0458
D Ga
0.002208 197.818
0.022023 61.9417
0.082915 24.7533
0.205399 10.6163
0.327835 4.68831
0.358575 2.03904
0.260599 0.849124
0.088355 0.320558
D Ga
1 0.320558
D Ga
1 0.1102
F Ga
1.0 0.3015

11.5 ECP associated with Peterson TZV basis set

Based on [47]
4 10
1
2 0.000000 1.000
2
2 370.273040 25.880361
2 9.190615 7.901295
4
2 99.144001 45.149190
2 198.295512 44.979981
2 28.445653 17.224251
2 56.949705 16.747329
6
2 -18.168797 51.968812 2 -27.380273 51.629117
2 -1.587022 15.241738
2 -2.516292 15.320193
2 0.083166 4.918589
2 0.202198 4.755103
2
2 -0.616990 10.762263
2 -3.138584 19.852939

11.6 Nitrogen

11.6.1 Basis set

Based on Def2-TZVP [15]
S N
0.21887984991E-03 19730.8006470
0.16960708803E-02 2957.8958745
0.87954603538E-02 673.22133595

0.35359382605E-01 190.68249494
0.11095789217 62.295441898
0.24982972552 22.654161182
S N
0.40623896148 8.9791477428
0.24338217176 3.6863002370
S N
1.0000000 0.84660076805
S N
1.0000000 0.33647133771
S N
1.0000000 0.13647653675
P N
0.55552416751E-02 49.200380510
0.38052379723E-01 11.346790537
0.14953671029 3.4273972411
0.34949305230 1.1785525134
P N
1.000000 0.41642204972
P N
1.0000000 0.14260826011
D N
1.0000000 1.65400000
D N
1.0000000 0.46900000

12 Appendix III - GaN luminescence lines and bands

Table 49: Luminescence lines and bands (eV) for GaN.Continued on next page. From Luminescence properties of defects in GaN, M.A.Reshchikov, H. Morkoc, J.Appl.Phys.97, **061301** (2005)

Maximum position (eV)	Doping	Comment
3.478	Undoped	
3.471	Undoped, Si	A few close lines
3.466	Undoped,Mg	Best FWHM <0.1meV
3.44-3.46	Undoped	Plethora of lines
3.455	Zn	A weaker peak at 3.39eV
3.45-3.46	Undoped	Correlates with inversion domains
3.41-3.42	Undoped	
3.397	Be	e-A type
3.387	Undoped	
3.38	Undoped	
3.38	Be	DAP type
3.37-3.38	Undoped	
3.375	Undoped	
3.364	Zn	
3.35-3.36	Undoped	
3.34	Undoped	
3.30-3.32	Undoped	
3.295	Undoped	
3.288	Undoped	
3.283	Undoped	
3.28	Undoped	e-A type
3.272	Zn	
3.27		DBE in cubic GaN
3.26	Undoped, S	DAP type
i 3.1-3.26	Mg	e-A and DAp
3.21-3.23	Undoped	
3.16		Shallow DAP in cubic GaN
3.08	Undoped	
3.08	C	In cubic GaN
3.0-3.05	C	Broad
2.9-3.0	Undoped, Fe	Broad, unstable intensity
2.9	P	Broad, with fine structure
2.88	Undoped	Broad, with fine structure
2.88	Zn	Broad, with fine structure
2.86	Undoped	
2.8	Undoped	
2.8	Cd	Broad, with fine structure
2.7-2.8	Mg	Broad, large shifts
2.6-2.8	Undoped	Broad, surface related
2.68	Undoped	
2.6	As	Broad, with fine structure

Table 50: Luminescence lines and bands (eV) for GaN, continued.From Luminescence properties of defects in GaN, M.A.Reshchikov, H. Morkoc, J.Appl.Phys.97, **061301** (2005)

Maximum position (eV)	Doping	Comment
2.6	Zn	Broad
2.56	Undoped	Broad
2.51	Undoped	Broad
2.5	Ca	Broad
2.4-2.5	Mg-O	Broad
2.48	Undoped	Broad
2.43	Hg	Broad
2.36	Undoped	Broad
2.2-2.3	Undoped, C	Broad
1.9-2.1	C	Broad, in cubic GaN
1.8-2.0	Undoped	Broad
1.85	Undoped	Broad
1.8	Zn	Broad
1.7-1.8	Mg	Broad
1.66	Undoped	Broad
1.64	C	Broad
1.3	(Fe)	Sharp
1.27	Mn	Broad
1.193	(Ti, Cr)?	Sharp
0.95	Undoped	Sharp, irradiation induced
0.85-0.88	Undoped	Sharp, irradiation induced